EdWorkingPaper No. 23-869

Information Frictions and Teacher Turnover

Kristen Shure
University of Pennsylvania

Zach Weingarten
University of Pennsylvania

Decentralized matching markets experience high rates of instability due to information frictions. This paper explores the role of these frictions in one of the most unstable markets in the United States, the labor market for first-year school teachers. We develop and estimate a dynamic model of labor mobility that considers non-pecuniary information frictions directly. We find that teachers overestimate the value of hidden amenities and their own preferences for teaching. Improving access to information improves stability by 12% and reduces between-school switching by 18%, but reduces teacher labor supply by over 5%. Compared to each tested alternative, including targeted wage premiums at hard-to-staff schools, bonuses that incentivize retention, and lowered tenure requirements, information revelation improves match quality most.

VERSION: October 2023

Suggested citation: Shure, Kristen, and Zach Weingarten. (2023). Information Frictions and Teacher Turnover. (EdWorkingPaper: 23-869). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/wdt4-pw47

Information Frictions and Teacher Turnover*

Kristen Shure
University of Pennsylvania

Zach Weingarten
University of Pennsylvania

October 4, 2023

Abstract

Decentralized matching markets experience high rates of instability due to information frictions. This paper explores the role of these frictions in one of the most unstable markets in the United States, the labor market for first-year school teachers. We develop and estimate a dynamic model of labor mobility that considers non-pecuniary information frictions directly. We find that teachers overestimate the value of hidden amenities and their own preferences for teaching. Improving access to information improves stability by 12% and reduces between-school switching by 18%, but reduces teacher labor supply by over 5%. Compared to each tested alternative, including targeted wage premiums at hard-to-staff schools, bonuses that incentivize retention, and lowered tenure requirements, information revelation improves match quality most.

JEL Classification: I20, I24, J45, J61, J63

Keywords: Information Frictions, Decentralized Markets, Teacher Turnover, Stable Match

*We are grateful for the continued guidance and feedback of Margaux Luflade and Petra Todd. We are also thankful for additional comments from Francesco Agostinelli, Jere Behrman, Hanming Fang, and Viviana Rodriguez, as well as other faculty and participants at the University of Pennsylvania Empirical Microeconomics lunch seminar. We finally thank Richard Ingersoll and Rebecca Maynard for assistance in data acquisition. This paper was supported by the National Science Foundation Graduate Research Fellowship Program and the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200035 to the University of Pennsylvania. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Contact information: Shure: University of Pennsylvania, klbeamer@upenn.edu. Weingarten: University of Pennsylvania, zachwein@upenn.edu.

1 Introduction

Decentralized two-sided matching markets often experience pairwise instability due to the existence of information frictions. Agents may be unaware of the full set of alternatives from which they can choose (e.g., Tsai and Honka, 2021), or they may have incomplete information about the alternatives they consider choosing (Abaluck and Compiani, 2020). Oftentimes, decision-makers learn important information *after* a match occurs, resulting in the subsequent dissolution of that match. In the marriage market, this information revelation can explain the decision of a couple to divorce (Li, Severinov and Xu, 2020); in the higher education market, the decision of a student to drop out of college (Arcidiacono et al., 2016); and in the labor market, the decision of a worker to leave a job (Baley, Figueiredo and Ulbricht, 2022). This paper explores the role of information frictions in decentralized matching markets by focusing on a particularly unstable market in the United States: the labor market for early career public school teachers.

The market for new teachers in the U.S. consistently produces some of the highest rates of college-educated worker turnover.[1] Job-level attrition for new teachers amounts to nearly twice that of new engineers and one-and-a-half times that of new nurses (Ingersoll et al., 2018). In fact, an estimated 44% of new teachers leave the profession during their first five years. Teachers additionally experience high rates of turnover within the profession, with 8% of all teachers switching schools annually (Gray and Taie, 2015).[2] Since the onset of the COVID-19 pandemic, these attrition rates have increased.[3] Pre-pandemic estimates suggest that teacher turnover in the U.S. is costly, with districts incurring upwards of $7 billion annually (Barnes, Crowe and Schaefer, 2007). Given the importance of public education funding in promoting student achievement, coupled with the fact that low-income districts experience the highest levels of turnover on average, these costs may also represent a large source of heightened education inequality for public school students (Simon and Johnson, 2015).

[1] Although we choose to focus on the United States due to data availability, high rates of teacher turnover are endemic to many developed countries with decentralized labor markets, like in Sweden (Lindqvist, Nordänger and Carlsson, 2014), the United Kingdom (Cooper and Alvarado, 2006), and Estonia (OECD, 2018).

[2] This rate is often much higher in urban areas. For instance, in Philadelphia, first-year teacher mobility exceeded 50% over the period 2010-2016 (Steinberg et al., 2018). Similarly, in 2017, 40% of all New York City teachers left low-performing schools after just two years (Zimmerman, 2017).

[3] Among members of the National Education Association, 55% (62%) of white (black) teachers reported plans to exit or retire early in January of 2022, compared to an average of 37% in August of 2021 (Walker, 2022).

In this paper, we estimate the effects of information frictions on teacher turnover by embedding distortions directly into a dynamic discrete choice model of early career teacher labor mobility. This structural approach allows us to test for the importance of unobserved frictions in a setting without information revelation interventions. Importantly, we use the estimated model to situate counterfactual information revelations against other policy-relevant counterfactual experiments like targeted wage bonuses, which we could not recover under a reduced-form framework alone. Standard models of labor choice include frictions in the form of probabilistic uncertainty, like termination or relocation, which do not necessarily depend on the market's degree of centralization. Our model extends this to include two additional sources of frictions that more directly reflect the decentralized structure of the teacher labor market. First, teachers cannot observe the full bundle of amenities that define a school at the time of receiving an offer. When a school offers them a teaching position, they directly observe the offered salary, the student proficiency level, the total proportion of non-white students, the commuting distance from home, and their average class size, but certain non-pecuniary characteristics—namely school climate and teacher workload—are learned only after they match. Teachers form rational expectations for school climate by relying on the local socioeconomic level as a noisy signal. Further, they anticipate teaching jobs to have the same workload as non-teaching jobs in the initial period. They subsequently update their beliefs each period given their actual workload experience.[4] In future periods, teachers know the amenities of their match school with perfect information but continue to face the same frictions at new schools that extend them a job offer. In the model, this channel reflects school-level (supply-side) learning.

Our second source of information frictions captures individual-level (demand-side) learning. New teachers may overestimate their level of commitment to the profession of teaching, where commitment is defined as the intention to remain a teacher in the long-run.[5] In the model, commitment captures preferences for being a teacher. Teachers learn their true level of commitment through on-the-job experience. While schools cannot observe teachers' *ex ante* belief about their level of commitment, we assume that teachers and the econometrician can. We con-

[4] We cannot observe similar information for non-teaching jobs and are limited to assume 40-hour work weeks in every non-teaching job.

[5] We could also estimate the fraction of teachers that underestimate their commitment level; however, making definitive statements about underestimated preferences for teaching are difficult to do with only three years of observation.

struct a binary measure of commitment using a combination of survey questions that ask new teachers about their long-run professional plans.[6] We make this distinction in teacher types to account for the fact that eventual counterfactual policies should not induce an identical reduction in turnover among teachers that plan to stay in the profession permanently and those that plan to exit shortly after beginning. We note too that new teachers may exit a school because the school itself has lower quality in amenities than they expected or because the experience of teaching was worse than expected. Disentangling these two is important to understand whether supply-side or demand-side improvements are first-order to mitigate turnover.

We estimate this model using a novel combination of data sources, including but not limited to the School and Staffing Survey (2007-2008), the Beginning Teacher Longitudinal Study (2007-2010), and the Baccalaureate and Beyond Longitudinal Study (1993-2003). There are several reasons these data are suitable for our analysis. First, the School and Staffing Survey provides detailed cross-sectional information on a nationally-representative sample of new teachers, in addition to their initial match school, their principals, and their district. This provides us with a holistic snapshot of teachers' initial experiences in the profession. Second, the Beginning Teacher Longitudinal Study follows this cohort of new teachers regardless of their labor decisions. This means that, unlike in administrative data, we observe the jobs that teachers hold even when they move states or change careers.[7] Importantly, this survey also asks teachers that leave their initial match school why they make the decisions they make, which gives us detailed information on which amenities they value most. Finally, the Baccalaureate and Beyond Longitudinal Study includes questions related to the number of schools a teacher applies to and receives an offer from, which is crucial since we only observe realized matches in our sample. This application data allows us to make inference on which schools comprise a teacher's choice set each year. We supplement other missing information important to our analysis using publicly available data sources, such as the American Community Survey and the Civil Rights Data Collection.

Our model estimates indicate that new teachers are largely uninformed about unobserved amenities. In the model, this level of noisiness induces teachers to down-weight their preferences for school climate, as measured by student misbehavior, toward zero. In doing so, they

[6] A detailed discussion of the construction of these types is presented in Section 3.

[7] For examples of papers that study teacher turnover with administrative data, see, e.g., Ronfeldt, Loeb and Wyckoff (2013), Adnot et al. (2017), and Sorensen and Ladd (2020).

anticipate student misbehavior to reflect the local socioeconomic average level; however, our descriptive results show that new teachers select into schools with disproportionately higher levels of student behavior than they should otherwise expect. Notably, this selection cannot be explained by differences in vacancies or supply-side hiring decisions, which we take as suggestive evidence in support of the role of information frictions in this market. At the same time, new teachers are overworked relative to alternative jobs. The average first-year teacher experiences a 55-hour work week, meaning that those who enter the profession expecting to work 40 hours will underestimate their true workload by over 37%. Lastly, we find that approximately 50% of new teachers overstate their commitment when entering the profession.

Using the estimated model, we simulate teachers' decisions in a labor market without information distortions. We find that providing teachers with increased access to information about the value of amenities results in large gains to match quality: three-year stable matches improve by 12%, relative to the baseline stable match rate of 63%; between-school switching decreases by 18% (baseline switching rate of 30%); and the fraction of former teachers at the end of the panel decreases by 34% (baseline fraction of 8%). The trade-off of these gains comes from the fact that providing this information reduces the overall supply of new teachers by almost 6%. Our next simulation reestablishes commitment as the recovered type, rather than the ones stated by teachers directly, to reflect learning induced by experience. This experiment improves stable matches by 4.7%, with no associated loss in labor supply.

We test for how well information revelation interventions perform against alternative policies proposed in the U.S. by situating these results against a battery of competing counterfactual experiments. We first introduce a universal salary premium of 10% at all schools. At an estimated cost of $44 billion, this policy produces effectively no change in the market for new teachers. Motivated by the policies outlined in Behrman et al. (2016) and Bobba et al. (2021), we next consider the effects of two different targeted policies, each designed to provide direct incentives for reducing turnover at a lower cost. We first provide a 10% retention bonus to teachers that stay at the same school year-to-year. This pay structure increases stable matches by 3% and decreases school switching by nearly 6%. In an effort to target schools that experience the highest levels of turnover, we consider another policy that provides the same bonus to teachers that work at

low-performing schools.[8] Doing so improves match quality at low-performing schools by 2%, but decreases matches at high-performing schools by nearly 1%. Overall, these results show that popular recommendations for policy designed to combat turnover miss a significant source of why this mobility occurs. School districts can avoid costly proposed solutions by instead increasing information access to new teachers. Through this, teachers can make more informed decisions, which in turn lead to more sustainable matches.

Our analysis relates to several strands of literature, primarily at the intersection of labor economics, the economics of education, and industrial organization. At a broad level, our results speak to the role of frictions in decentralized matching markets. Although many papers explore the stability of matches under centralized mechanisms, fewer papers consider the extent to which these frictions cause instability under decentralization.[9] On the theoretical front, Niederle and Yariv (2007) show that stable matches can occur in decentralized markets when both sides have complete information. Similarly, Dai and Jordan (2021) derive optimal strategies for two agents in a matching market absent a central planner and show that learning may cause eventual stability. Some empirical work explores the effects of asymmetric information on labor market outcomes (Baldwin, 1991; Golan, 2005; Bartanen and Kwok, 2022) and consumer search (Buchholz, 2022), but unlike our work, these papers do not consider the explicit effect of eliminating the friction on match quality.

Our findings also relate to a smaller strand of literature on labor search with a focus on teachers. Standard models of labor search seek to explain firm wage-setting behavior and worker-firm match patterns under an assumption of sequential job offer flows that workers choose to either accept or reject (Eckstein and Wolpin, 1990; Van den Berg and Ridder, 1998). Several characteristics of the teacher labor market make this typical model a poor fit. First, schools do not set wages directly; instead, they are legally required to take salary schedules handed down by the district or state as given.[10] As a result, public schools do not compete with one another through wage-setting. Second, teachers cannot bargain for higher salaries due to the inflexibility of these wage schedules. In most states, the only factor that impacts salary is the number of years of

[8] We define a low-performing school as one whose average student proficiency level is in the bottom 25% of its state.

[9] See Agarwal (2015) for a discussion and application on many-to-one matches and Abdulkadiroğlu et al. (2020) for theoretical considerations of one-to-one matches.

[10] Like private schools, the salaries at charter schools are often much lower than at public schools. Additionally, we observe a negligible fraction of teachers that switch into charters.

teaching experience an applicant has. Third, because schools need to fill vacancies for the school year, hiring decisions happen universally in the spring and summer for upcoming spots in the fall. There is no on-going search process with a constant potential flow of job offers for teachers. As a result, we model teachers' labor market decisions in a given period as one in which they choose among a finite set of simultaneously occurring offers.

Related to the previous strand of work, ours contributes to a body of research that models teachers' labor decisions in a discrete choice dynamic programming (DCDP) framework.[11] Stinebrickner (2001) developed the first discrete choice dynamic programming problem that focused on teachers' labor choices as a result of wage changes.[12] Wiswall (2007) expands this model to include teachers' licensing decisions in the context of mitigating turnover. Most recently, Behrman et al. (2016) introduces the subsequent decision teachers face to work in the public or private education sector, while Ederer (2023) develops a dynamic model of teacher sorting to show that the centralized Peruvian teacher labor market operates as a spatial job ladder. Each of these papers considers the full labor force of teachers. We instead focus on early career teachers to highlight the determinants of turnover associated with the first few years of the job. We are also the first to directly embed information frictions into this dynamic model, which gives a larger consideration to non-pecuniary amenities and their impact on teacher mobility.

In addition to those that rely on DCDP techniques, several recent papers develop structural models of teacher-school matches. Biasi, Fu and Stromme (2021) estimate a static general equilibrium in which districts strategically choose policies to hire their most preferred teachers. Their model features wage rigidity as an important market friction but omits information frictions. Similarly, Bates et al. (2022) use application data to estimate the preferences of both teachers and principals in a labor search model. Other work, like our own, considers the formation of teacher-school matches in a partial equilibrium framework (Dolton and van der Klaauw, 1999; Stinebrickner, 2002; Feng, 2009). One notable limitation to nearly all of these studies is their use of administrative data, which vastly limits the possible scope to analyze attrition, since the data

[11] In a similar vein, Johnston (2020) develops an experiment to recover teachers' willingness-to-pay for various job attributes, finding teachers are highly elastic to certain non-pecuniary factors like commuting time and principal quality.

[12] This is the first published work. There exist unpublished manuscripts which predate Stinebrickner (2001) that feature discrete choice teacher mobility.

do not allow researchers to differentiate between competing explanations for sample attrition.[13] We circumvent this issue by utilizing rich survey data that follows teachers regardless of their labor decision. Importantly, this survey data includes questions that directly ask former teachers to explain why they left, and for teachers that switch schools to directly compare their new school to their old one. The specificity of these questions significantly reduces the burden we face to infer why teachers make the choices they do.

Lastly, we contribute to the large economics of education literature on teacher turnover. A subset of this strand focuses on the student-level effects of teacher attrition and vacancy, usually finding large, negative and lasting impacts (Hanushek and Rivkin, 2010; Chetty, Friedman and Rockoff, 2014; Hanushek, Rivkin and Schiman, 2016). Another collection of work looks to explain why teachers leave and what incentives they respond to, often analyzing policies which give teachers more pecuniary benefits, either in the form of sweeping salary increases or targeted vouchers (Behrman et al., 2016; Bobba et al., 2021; Morgan et al., 2023). Less often, researchers consider the role of non-pecuniary school amenities to explain teacher turnover (Falch and Strøm, 2005; Feng, 2009; Bonhomme, Jolivet and Leuven, 2016). Our work relates most to this final group of papers on non-pecuniary amenities, although ours is the first to consider the role of information frictions as an explanation.

The rest of the paper is organized as follows: Section 2 describes the data sources. Section 3 presents several motivating descriptive analyses. Section 4 presents the dynamic structural model of teacher preferences. Section 5 outlines the estimation strategy and sources of identification. Section 6 provides the corresponding estimates and validates the model. Section 7 outlines the counterfactual experiments and presents the results. Finally, Section 8 concludes the paper.

2 Data

2.1 Primary Data Sources

We use a combination of several restricted-access data sources provided by the National Center for Education Statistics (NCES), along with several public data sources, to construct our final

[13] For example, a Chicago teacher who remains in teaching but moves to another district, and a teacher who takes a non-teaching job in Chicago, are indistinguishable from the perspective of a researcher with data on the district.

analysis sample. Our primary data sources include the School and Staffing Survey (2007-2008), the Beginning Teacher Longitudinal Study (2007-2010), and the Baccalaureate and Beyond Longitudinal Study (1993-2003). The School and Staffing Survey includes four surveys designed to understand the holistic state of public education: one answered by teachers, one by principals, one that provides school-level information, and another that provides district-level information. The teacher questionnaire includes rich information on teachers' demographics, wages, education, certification and training, professional development, and opinions on school climate and performance.

The NCES used the responses to the School and Staffing Survey to identify first-year teachers and subsequently construct the Beginning Teacher Longitudinal Study (BTLS). The BTLS surveys these first-year teachers for an additional four years, designed to measure the attrition rate of teachers and qualitatively explain why many teachers exit so early in their careers. This survey has two attractive qualities for the purposes of this paper: first, teachers are asked to choose among a list of reasons why they decide to either switch schools or leave teaching altogether between years; second, unlike administrative data on public school teachers, respondents remain in the survey even when they switch to private schools, move states, or take on non-teaching jobs. A combination of factors lead us to restrict the analysis to the first three years of the survey, which still allows us to capture the majority of movement during a teacher's early career period.[14]

Common to research focused on decentralized two-sided matching markets, we only observe the realized match between each teacher and their selected school in each year. We do not directly observe the schools a teacher rejects, nor does the BTLS ask any questions related to application behavior on the part of teachers. To circumvent this issue and provide structure to the choice sets of teachers, we rely on job search data from the Baccalaureate and Beyond Longitudinal Study (B&B). The B&B asks recent college graduates who enter the teaching profession how many schools they apply to and how many schools offer them a full-time position. We use this information to impute the number of offers new teachers receive in the BTLS, using a procedure outlined in Section 4.

[14] The BTLS switched from in-person interviews and phone call interviews to a strictly online survey during the third year and beyond. As a result, response rates rapidly diminish. Nearly 25% of teachers do not respond in the fifth year. At the same time, the number of questions and specificity of the survey questions decreases substantially over time.

2.2 Supplementary Data Sources

We utilize several sources of publicly available data to supplement information unavailable in our primary data. First, we use a collection of data merged together by the Urban Institute, which includes the Common Core of Data of America's Public Schools and the Civil Rights Data Collection. This provides us with school-level information on the racial composition of each school. We also obtain math and reading proficiency levels for each school, which we average and then standardize within each state and school level.[15] We match these data to each school in the BTLS.[16] These data also contain exact coordinates for each school, which we use to compute the geodistance between a teacher's match school and their imputed home location.[17] To estimate travel times and salary offers for teaching and non-teaching jobs, we use the American Community Survey (ACS), which we supplement with the 2008-2012 ACS 5-year estimate to classify each U.S. ZIP code as either high or low socioeconomic status (SES). Lastly, we use annual reports published by the National Council on Teacher Quality (NCTQ) to track changes in state-level education policies.[18] This provides us with details related to, e.g., the number of years teachers must work before earning tenure.

2.3 Sample Restrictions

The BTLS initially contains about 2,000 first-year teachers. We make several restrictions to the sample in order to mitigate issues related to missing data and outliers. We first restrict the sample to those who, at most, have earned master's degrees, since those with PhDs typically intend to become principals and are therefore not representative of the average early-career teacher. We next remove first-year teachers close to retirement age. Specifically, we drop anyone above the age of 58 in the first year, although very few teachers in the data are above the age of 50, so the

[15] The United States does not administer any universal standardized tests in primary or secondary school. Instead, each state independently administers a combination of exams at various grade levels, usually beginning in 3rd grade. As a result, we consider within-state standardizations of proficiency levels. We do this separately for elementary, middle, and high schools to account for differences in achievement under the assumption that e.g. elementary school teachers do not consider local high school quality when deciding where to work.

[16] Although we do not observe the set of schools each teacher rejects, we simulate offer sets each period based on procedures outlined in Appendix C.

[17] Teachers' home addresses are censored, so we instead set their address to the latitude and longitude of the ZIP centroid for their match school.

[18] These reports form a series of publications known as the *State Teacher Policy Yearbook*. Each report can be found on the NCTQ website here. After 2010, these reports became biannual (every other year).

exact cutoff does not impact our analysis. We also remove the few teachers who pass away during the period, as well as those who did not answer survey questions indicating their commitment to the profession, which we use to define a teacher's type, as discussed in the next section. Finally, because we focus on the first three years after a teacher begins working in the profession, we remove teachers that do not respond to the survey in the third year.[19] After implementing all of these restrictions, we retain a final analysis sample of 1,740 first-year teachers.[20],[21]

3 Descriptive Statistics and Reduced-Form Analysis

3.1 Descriptive Statistics

Table 1 provides summary statistics for the sample during the first year of teaching. Columns 1 and 2 give the mean and standard deviation for the full sample of new teachers. Panel A displays the spread of demographic characteristics. The majority of new teachers are white (79%) and female (76%). Effectively all new teachers have a college degree, but less than 20% have a master's degree and only two-thirds enter with state certification. Nearly 75% of the sample have a degree in education, which we define as either being an explicit education degree (e.g., Early Childhood Education) or as one granted by their college's school of education. Lastly, the average new teacher in our sample is 29 years old.

Panel B displays summary statistics for various measures of teachers' professional careers. These measures include their base salary[22] and their attachment to schools, as indicated by additional responsibilities. New teachers earn low salaries, with an average base of approximately $36,000 and a standard deviation of less than $8,000.[23] Nearly half of teachers take on additional responsibilities (coaching a team or sponsoring a club) and over half of the sample belong to a

[19] We test for whether these missing data occur at random and display the results in Table A1. We find no evidence that non-response in year three is significantly predicted by observed characteristics of teachers, so we proceed without these teachers in our analysis.

[20] Per guidelines outlined by the National Center for Education Statistics (NCES), we round the sample size to the nearest 10 to preserve anonymity.

[21] The final sample is representative of the teacher labor force in the United States. The relative spread of initial match schools closely matches population patterns and contains all 50 states, as well as the District of Columbia.

[22] Although the BTLS includes information on benefits, we exclude these from our measure of salary since we only obtain salary information for non-teaching jobs.

[23] Differences in state salaries drive this heterogeneity. For example, the average starting salary of North Dakota teachers in the sample is $26,000 while the average in California is about $45,000.

Table 1: Descriptive Statistics for Public School Teachers, by Commitment Level

	Full Sample		Committed		Uncommitted		Difference in Means
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	by Commitment Type
Panel A: Demographics							
Female	0.758	0.428	0.764	0.425	0.746	0.436	0.018
Black	0.069	0.253	0.065	0.247	0.078	0.268	-0.012
Hispanic	0.105	0.306	0.114	0.318	0.085	0.278	0.029*
White	0.791	0.407	0.787	0.410	0.800	0.401	-0.013
Bachelor's	0.988	0.107	0.990	0.101	0.986	0.119	0.004
Master's	0.184	0.388	0.172	0.377	0.212	0.409	-0.041**
State Certified	0.667	0.471	0.679	0.467	0.642	0.480	0.037
Degree in Education	0.724	0.447	0.755	0.430	0.653	0.477	0.102***
Age	28.98	8.419	29.26	8.33	28.36	8.60	0.899**
Panel B: Career Attributes							
Union	0.626	0.484	0.642	0.480	0.589	0.492	0.052**
Sports Coach	0.176	0.381	0.194	0.396	0.134	0.341	0.061***
Club Sponsor	0.282	0.450	0.281	0.450	0.283	0.451	-0.002
Base Salary ($)	36,133	7,654	36,260	7,087	35,847	8,793	413.78
Uncompensated Payments ($)	197.87	368.61	224.12	406.91	137.21	249.26	86.91***
Non-Teaching Experience	2.434	5.034	2.483	5.135	2.324	4.804	0.158
Weekly Hours	55.06	10.10	55.30	9.856	54.52	10.61	0.785
Panel C: School Characteristics							
Public	1.000	0.000	1.000	0.000	1.000	0.000	—
Elementary	0.454	0.498	0.465	0.499	0.430	0.495	0.035
Middle	0.177	0.382	0.196	0.397	0.133	0.340	0.063***
High	0.304	0.460	0.275	0.447	0.370	0.483	-0.096***
Rural	0.440	0.497	0.464	0.499	0.388	0.488	0.076**
Suburban	0.308	0.462	0.305	0.461	0.316	0.465	-0.011
Urban	0.251	0.434	0.231	0.422	0.297	0.457	-0.066***
Charter	0.046	0.210	0.038	0.192	0.063	0.243	-0.025**
Fraction Non-White Teachers	0.183	0.237	0.169	0.217	0.216	0.276	-0.047**
Fraction Non-White Students	0.495	0.350	0.483	0.345	0.521	0.361	-0.038**
Fraction FRPL Students	0.472	0.285	0.471	0.281	0.474	0.291	-0.002
Class Size	21.22	8.477	21.03	7.641	21.65	10.09	-0.615
Observations	1,740		1,140		600		—

NOTES: We report weighted sample averages from the BTLS sample. We then divide this sample by commitment type—committed and non-committed—and report their respective averages as well. Column 7 provides results from weighted bivariate regressions which test for differences across commitment level with robust standard errors. We weight by the final survey weight provided by NCES and round sample sizes to the nearest 10. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

union. Compared to a standard 40-hour work week, new public school teachers are greatly overworked. On average, teachers spend 55 hours a week performing school-related tasks, including

teaching, preparing lesson plans, and volunteering.[24]

Lastly, we display information on the types of schools teachers match with in Panel C. Notably, every teacher in the BTLS begins at a public school. Nearly half work at elementary schools while 30% work at high schools. Teachers tend to work in schools with a relatively high proportion of non-white students, but a relatively low proportion of non-white teachers. Somewhat similarly, most teachers work at schools with a high proportion of students entitled to free or reduced-price lunch, indicating low levels of household income. Class size varies across state and school level, with a typical teacher having a class size of about 21 students.

Table 1 additionally displays these statistics by teachers' commitment type. In this paper, we define commitment as an *ex ante* measure of whether a teacher plans to remain in teaching in the long-run. We identify this using a combination of survey questions in the BTLS. The first asks, *If you could go back to your college days and start over again, would you become a teacher or not?* The second question asks, *How long do you plan to remain in teaching?* We report the responses to these two questions in Figure B1. Under the first question, we denote a teacher as committed if they answered "certainly" or "probably" (uncommitted otherwise). For the second, we take any answer indicating plans to stay in teaching for the long-run as an indication of commitment. We use the rule that *both* responses must indicate commitment; through this, we find that 68% of new teachers are committed.[25]

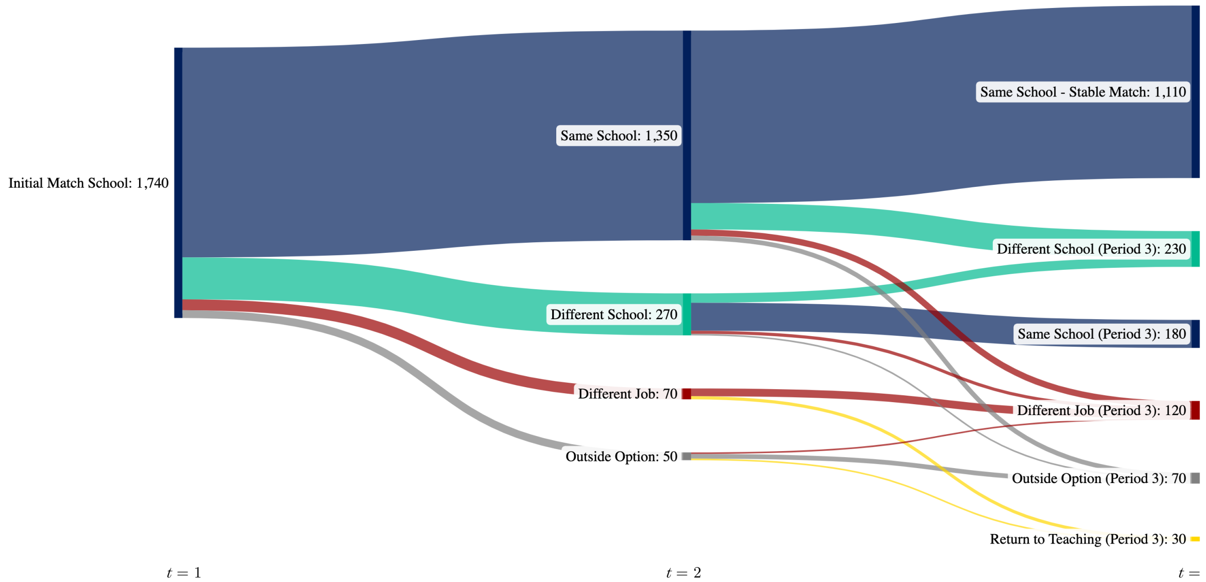
One potential downside to relying on these questions comes from the fact that NCES surveys teachers during their first year of teaching. As a result, we may capture lower proportions of committed teachers if some respondents already had poor experiences in their initial match school before the time of survey. However, we test for the propensity of "committed" responses across survey period and find no statistically significant differences.[26] In other words, we find no evidence to suggest that teachers surveyed towards the end of the school year give less committed

[24] We scale each teacher's reported average to a 40-hour work week to adjust for differences in the required number of hours teachers work to receive their base pay. For example, a teacher that reports 35 hours of work on average at a school that compensates them for 30 hours is adjusted to work 47 hours a week for comparability.

[25] Relaxing this assumption can yield estimates as high as 90%, which make estimating separate parameters by teacher type infeasible. Further, given the level of teacher movement observed, 90% seems to substantially overestimate the proportion of committed teachers. The covariance between these two questions' measure of commitment is 0.83, suggesting a strong positive relationship in the answers to both.

[26] The 95% confidence interval obtained from a test of the difference in propensity to be committed depending on being surveyed in the fall versus the spring semester is $(-0.059, 0.032)$.

Figure 1: Labor Decision Flows Among Early Career Teachers



NOTES: The figure above tracks early career teachers' labor market decisions during their first three years in the profession. Numbers reported at each node refer to weighted counts of observations using analysis weights provided by NCES, which are then rounded to the nearest 10 to preserve anonymity. SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

responses than teachers surveyed towards the beginning of the school year. With this in mind, we proceed under the assumption that our measure of commitment is accurate and informative.

In the final column of Table 1, we present results from regressions that test for differences between committed and uncommitted teachers. We find no differences in commitment on the basis of gender or race, but find some evidence to suggest that those with a master's degree tend to be less committed to teaching, as well as those without an education degree. Committed teachers more often volunteer for additional responsibilities within their school and pay for supplies out-of-pocket. Evidence from Panel C suggests that committed teachers tend to work in slightly different schools, like middle schools or those located in rural areas. We formally test whether observed characteristics of teachers predict their commitment type (presented in Table A2) and find little evidence to suggest that commitment is predictable at the time a teacher applies to schools.

Figure 1 displays the mobility of new teachers across the three year panel. In order of frequency, teachers most often stay at their current school, move to a different school, remain in the labor force but leave teaching, and leave the labor force altogether (denoted by the outside

option). After one year, 23% of teachers leave their initial match school; after two years, fewer than 64% of teachers remain at their first school. Between-school switching accounts for 68% (56%) of total turnover in the second (third) year.

Surprisingly, we find no substantial differences in the propensity to switch schools in the third year, regardless of the switching decision in the second year. About 14% of teachers who stay at their match school for two years switch schools in the third year, while 18% of previous switchers actually switch again in the final year. Although these results point to the total level of mobility, they provide no indication for why teachers make the decisions they make. To understand this, we investigate survey responses among those that leave their match school in either period.

3.2 Predictors of Mobility

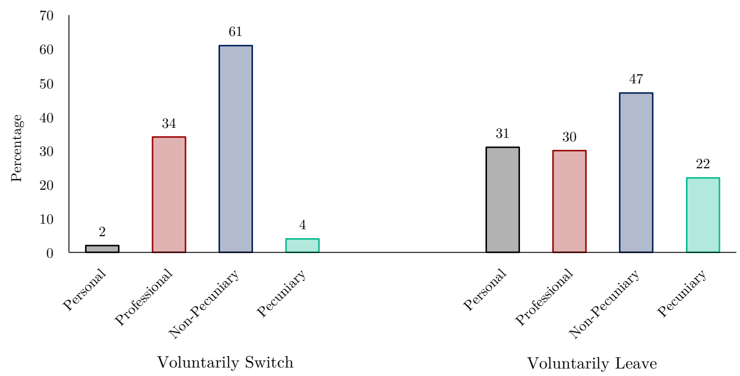
Figure 2 shows the frequency of responses given by voluntary switchers and leavers when asked for the most important reason impacting their decision to leave their initial match school. Figure 2a provides the distribution in the second year while Figure 2b gives the responses in the third year. In both panels of the figure, we condition on voluntary movement to more directly capture the preferences that drive mobility.[27] For simplicity, we aggregate heterogeneous responses into one of four groups: personal, professional, non-pecuniary, and pecuniary.[28]

What teachers value impacts their decision greatly. In the second year, the vast majority of teachers that switch schools do so because of the non-pecuniary amenities of their match school, above all. These amenities hold relatively less importance to teachers that exit the profession, with about half as many citing this as their primary reason for leaving. In the third year, pecuniary amenities become an even less important factor. Additionally, non-pecuniary amenities become more important and professional reasons become slightly less important. We observe two main facts from this figure. First, teachers rarely make labor changes on the basis of salary or benefits, even when they move to schools that pay more, as evident in Figure B4. Second, non-pecuniary

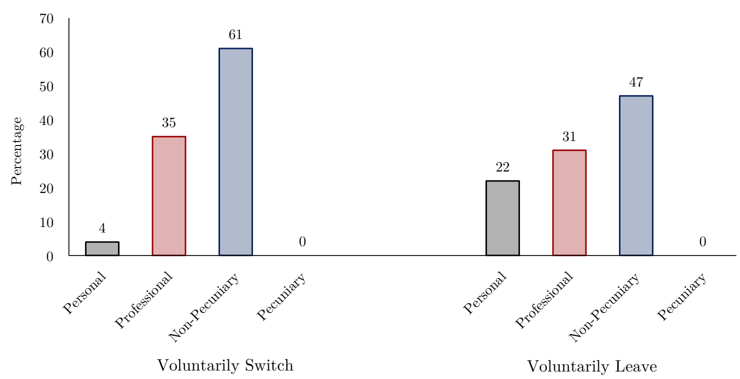
[27] In other words, we drop fired teachers only from this figure since their preferences do not drive their initial match dissolution. We do not make this distinction elsewhere in the paper, meaning these teachers still comprise the sample from which we estimate preference parameters. Fired teachers account for approximately 25% of movers.

[28] For clarity, we provide some of the many examples that fall under each category: "personal" refers to things like health shocks or family death, "professional" refers to career-specific attributes like grade level assignment and the desire to be a teacher, "non-pecuniary" refers to school characteristics like workload, administrative support, and student behavior, and "pecuniary" strictly refers to salary and benefits.

Figure 2: The Most Important Factor for Voluntary Mobility Decisions



(a) Responses in Year 2



(b) Responses in Year 3

Notes: The above figure displays the most important reason switchers and leavers attribute to their choice to leave their original school. The left-hand side displays responses for those that remain in teaching but switch schools. The right-hand side displays the same but for those that leave the profession. In both instances, we consider only voluntary movers, i.e., those not fired from their school. Each panel only considers *new* movers, meaning the preferences of those who switch in period 2 are not accounted for in panel (b). In accordance with NCES guidelines, we censor fractional shares to 0. Source: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

amenities increasingly influence teachers' decisions to leave the profession. In Figure B5, we provide detailed information on how second period switchers compare characteristics of their new school to their original one. Together, these three figures highlight the relative importance of amenities in teachers' mobility decisions and demonstrate why research on teacher mobility should give heightened attention to these characteristics.

We tie these findings together by exploring the role of commitment in the mobility decisions. In Table 2, we present the results from several regressions which estimate the impact of commitment status on the eventual decision to change schools or change job sectors. Columns 1 and 2 display the effects that being a committed teacher has on the propensity to switch from the initial

Table 2: Regression Results for Likelihood to Switch or Exit

	Ever Switch Schools		End as Non-Teacher	
	(1)	(2)	(3)	(4)
Committed	0.077**	0.084**	−0.083***	−0.074***
	(0.034)	(0.034)	(0.023)	(0.022)
Female	0.055	0.048	0.026	0.033
	(0.044)	(0.042)	(0.027)	(0.028)
Non-White	0.103	0.073	−0.067**	−0.087**
	(0.124)	(0.114)	(0.033)	(0.041)
School Fixed Effects	✓	✓	✓	✓
State Fixed Effects	✗	✓	✗	✓
Observations	1,740	1,740	1,740	1,740

NOTES: School fixed effects refer to the school level (elementary, middle, or high) effects, subject area effects, and urbanicity effects. In addition to the listed covariates, each regression controls for age, the proportion of non-white students at a school, the interactions between female, non-white, and non-white students, distance from home, and the number of IEP and LEP students a teacher has. We report heteroskedastic-robust standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

match school. Under each specification, we find a significant increase in the likelihood to switch among committed types. Consistent with our definition, these teachers more often search for a better match due to the fact that they intend to remain in the profession in the long-run. Again consistent with our definition of commitment, we find that these types of teachers are significantly less likely to exit the profession in the short-run (Columns 3 and 4). These estimates point to the fact that teachers' beliefs about their type meaningfully predict their choices.

3.3 A Competing Theory of Turnover

We lastly consider a competing explanation for the high levels of first-year teacher mobility: the job ladder model.[29,30] If new teachers experience a labor market which follows a "job ladder", then they systematically receive offers only from low quality or undesirable schools. New teach-

[29] This has been found to be the case outside of the United States, e.g. in Ederer (2023).

[30] Another competing explanation is that the majority of new teachers belong to Teach For America (TFA), an organization which situates temporary first-time teachers in often low-performing schools with high vacancy rates. We do not observe whether a teacher belongs to TFA; however, we do not believe this drives high rates of turnover since only about 5,000 teachers belong to any given TFA cohort and approximately 60% of TFA teachers stay at schools beyond their required two years of service (Coffman et al., 2019).

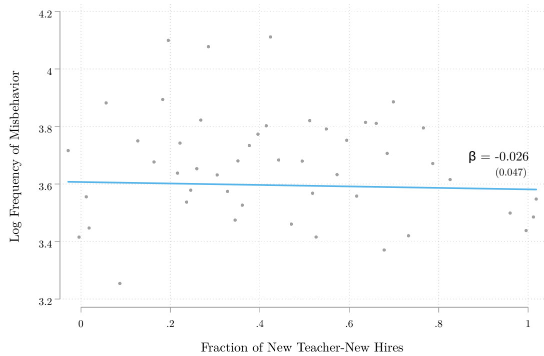
ers would then spend the first few years of their career gaining experience in schools with the highest levels of turnover until vacancies open in more desirable schools. This could directly contradict our explanation of turnover by information frictions if instead high rates of turnover are a mechanical response to both the market structure and the constrained choice sets that new teachers face. In other words, teachers leave schools not because they learn about unknown attributes, but because offers from better schools eventually become available. Broadly, we find this explanation unlikely due to the fact that 86% of public schools in 2008 had vacancies and, among schools with at least one vacancy, 80% hired new teachers. In an effort to more rigorously understand whether the job ladder can explain movement in our sample, we directly investigate the differences in public schools by the share of new teachers they hire, as well as the differences between schools in either low or high SES areas.[31] We present these results in both Figure 3 and Table 3.

Figure 3 plots OLS estimates for different measures of school desirability by the fraction of new hires that are first-year teachers. Each specification controls for the relative urbanicity of a school, although these results are not sensitive to whether we instead control for, e.g., the state a school resides in. Results consistent with the job ladder theory of increasing quality over teacher experience would show that less desirable schools hire a greater fraction of new teachers relative to schools with higher demanded amenities; instead, we find little evidence that any relationship exists at all. Figure 3a shows the estimates for our main amenity of interest, the level of misbehavior incurred at a school. For this, we find no statistical difference between schools that hire zero first-year teachers to fill their vacancies and schools that exclusively hire first-year teachers.

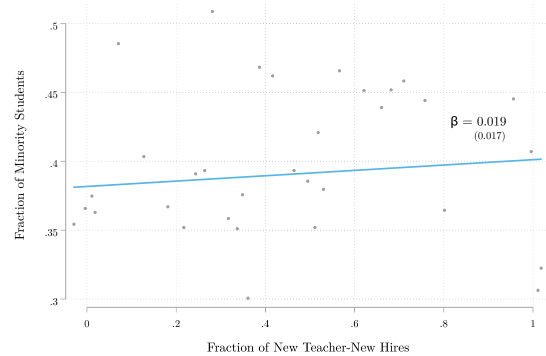
Other student level characteristics, like the fraction of minority students (Figure 3b) and the fraction of students with an Individualized Education Program, or IEP, (Figure 3c), also show no significant relationship. We lastly find a small statistical difference in the relative probability of working in a high SES area, shown in Figure 3d, although the actual magnitude remains small at about 7 percentage points. This means that a school that fills its total share of vacancies with no new teachers has only a 7% higher likelihood of being located in a high SES area than a school

[31] We use the ACS to obtain the median income in the US during the period and say a ZIP is low (high) SES if the average income in that ZIP is below (above) the median of $51,000.

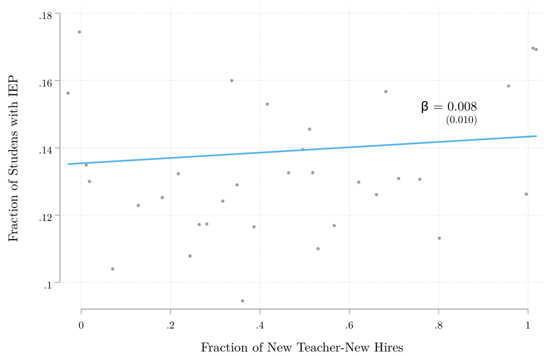
Figure 3: Public School Attributes by Fraction of Vacancies Filled by New Teachers



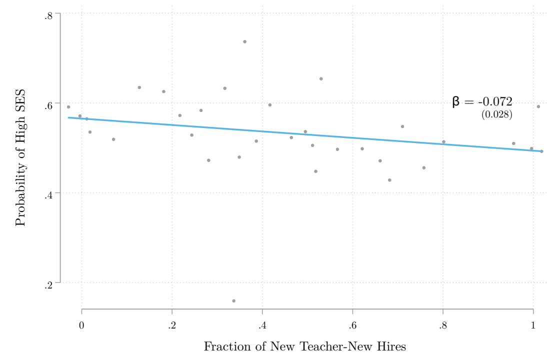
(a) Misbehavior



(b) Minority Students



(c) Students with IEPS



(d) SES Level

NOTES: Each panel plots a fitted linear regression model of the listed outcome on the proportion of vacancies filled by first-year teachers. In each specification, we control for the public school's urbanicity but make no other adjustments. We obtain comparable estimates if we instead control for the U.S. state a school resides in. Coefficients noted above each line refer to the estimate obtained from the OLS model, with the corresponding heteroskedasticity-robust standard error in parentheses below. SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public School Questionnaire", "Public School Teacher Questionnaire", 2007-2008.

that fills all of its vacancies with new teachers. Together, we take these results to suggest that no notable gradient exists between the quality of a school and the fraction of vacancies allocated to new teachers.

We conclude this section by showing evidence that new teachers select their initial match school uninformed about student behavior. Table 3 first displays the difference between low and high SES public schools and then shows how first-year teachers select schools in these same areas. Evident from the first two columns, public schools largely differ by SES level. On average, low SES schools have higher rates of misbehavior, a larger fraction of students with IEPs, and more often have a student body comprised of majority-minority students. Despite this, first-year teachers

Table 3: Heterogeneity in Public Schools by Share of New Teachers and SES Level

	All Public Schools		Public Schools Chosen by First-Year Teachers	
	Low SES	High SES	Low SES	High SES
Log Behavioral Infractions	3.617	3.519	3.603	3.698
	(0.863)	(0.361)	(0.727)	(0.828)
Students with IEP	0.150	0.140	0.120	0.155
	(0.158)	(0.159)	(0.173)	(0.250)
Majority Minority Students	0.459	0.321	0.645	0.248
	(0.364)	(0.295)	(0.479)	(0.430)

NOTES: The above table considers the socioeconomic status (SES) level of initial match schools. Panel A shows the discrepancy between the distribution of log behavioral infractions in the full sample of public schools in the School and Staffing Survey (SASS) [N=7,190] and the new teachers' selected schools. Panel B documents how new teachers sort into schools located in different SES areas. SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public School Questionnaire", "Public School Teacher Questionnaire", 2007-2008.

tend to work in schools which are uncharacteristic of the population average, especially for high SES schools. In particular, new teachers working in these areas select schools with an observably lower fraction of minority students but a higher level of misbehavior. These schools additionally have a higher rate of behavioral infractions compared to the set of low SES schools. We take these results as suggestive evidence that first-year teachers select schools without knowing the true level of student misbehavior.

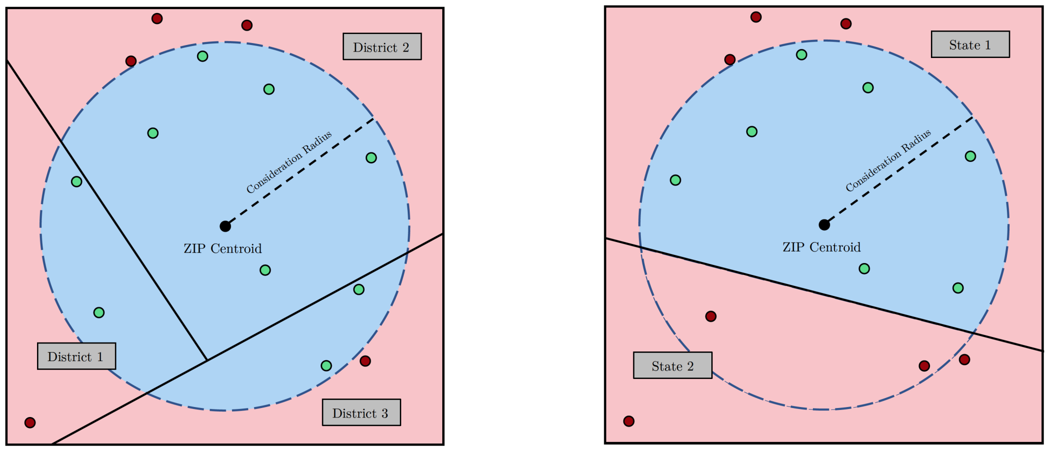
4 Model

4.1 Environment

We consider the market for early career teachers in the United States for a given year. This market is defined by its supply of new teachers, I, and its stock of schools with vacancies, J. Each teacher $i \in I$ applies to a set of public schools, forming their application set, $\mathcal{A}_i$. Of these schools, at least one school j extends teacher i an offer, which forms their offer set, $\mathcal{O}_i \subseteq \mathcal{A}_i$.[32] As in Agarwal and Somaini (2022), we say teacher i and school j match if school $j \in \mathcal{O}_i$ and teacher i chooses

[32] In general, this need not be true; however, our data includes only those who become teachers and so for our purposes the offer set size is positive.

Figure 4: Application Consideration Set for Early Career Teacher



(a) Home Near District Boundaries

(b) Home Near State Boundaries

Notes: Each colored dot represents a public school j close to teacher i's home. Those shaded in green have a distance d_{ij} less than or equal to the consideration radius r_i. Conversely, those shaded in red are defined such that $d_{ij} > r_i$ or j falls outside of the home state. Panel (a) illustrates the fact that teachers do not condition on their home district when considering schools to work at. Panel (b) highlights the assumption that teachers only consider schools in their home state.

it. Matches occur prior to all information being revealed to either side of the market.

Each school is defined by seven characteristics, two of which remain unknown to teachers up until the point of matching. We assume that teachers know (1) the exact salary offered, (2) the proficiency level of students at the school, (3) the proportion of non-white students at the school, (4) the distance from home they will need to commute, and (5) their class size. These assumptions are reasonable given that job offers explicitly outline salaries (as well as their schedules) and teachers interview for these positions in-person. Further, information like proficiency levels and teacher-student ratios is publicly available to teachers who search for these details on school quality websites.[33] Schools also have a climate as measured by their level of student behavioral infractions. The true value of a school's climate is not known by teachers until they work at that school. Furthermore, teachers do not know the true workload required at a school.

Although we do not explicitly model the decision of a teacher to apply to a given school, we make some formal assumptions to restrict their choice set.[34] We use data from the B&B to predict the number of offers a first-time teacher with a given set of observed characteristics receives. This

[33] One example includes greatschools.org.

[34] Agarwal and Somaini (2022) detail how to estimate parameters in the presence of latent choice sets. Given our lack of school-side market data, we rely on this work to instead develop a structure to the choice set formation.

prediction gives us the size of a teacher's offer set, $|\mathcal{O}_i|$.[35] We assign a consideration radius, r_i, to each teacher i, which denotes their maximum willingness to travel from home to work.[36] Lam, Broderick and Toor (2018) detail how travel times vary by rurality level. We fix the consideration radius to 45 travel minutes, and so teachers living in urban (rural) areas consider schools at most 19.25 (27.75) miles from their home.[37,38] Figure 4 provides stylized illustrations for the ways teachers incorporate these consideration radii into district and state boundaries.[39]

4.2 Utility and Information

Teacher i receives utility $u_{ij\ell t}$ from working at school j with an associated SES level $\ell \in \{L, H\}$ in year t, which takes the form

$$u_{ij\ell t} = \alpha_i \ln w_{j\ell t} + \beta_i \ln b_{j\ell} + \zeta z_{j\ell} + \gamma_i m_{j\ell} + \delta d_{ij\ell t} + \lambda(40 - h_{ij\ell t}) + \psi_1 \bar{n}_{ij\ell t} + \psi_2 \bar{n}^2_{ij\ell t} + \phi_i + \varepsilon_{ij\ell t}. \quad (1)$$

In Equation 1, $\ln w_{j\ell t}$ denotes the offered (log) salary, $\ln b_{j\ell}$ the frequency of behavioral infractions, $z_{j\ell}$ the average level of student proficiency, $m_{j\ell}$ an indicator for whether non-white students comprise the majority of a school's student body, $d_{ij\ell t}$ the geodistance in miles between the school and a teacher's home, $h_{ij\ell t}$ the number of hours worked in an average week, and $\bar{n}_{ij\ell t}$ their average class size.[40] We assume teachers derive disutility from travel.[41] The term $40 - h_{ij\ell t}$ captures deviations from the standard work week.[42] We assume teachers have positive, marginally decreasing preferences for class size as captured by the quadratic form with $\psi_1 > 0$ and $\psi_2 < 0$.

The preference parameters indexed by i flexibly allow for heterogeneity at the margin of

[35] Given our estimates, we assume all teachers have an offer set size of two schools.

[36] Due to the fact that NCES censors teachers' homes locations, we impute these value as the ZIP centroid of the initial match school.

[37] Teachers living in suburban areas have a consideration radius of 21 miles.

[38] Santelli and Grissom (2022) estimate an average commute time for teachers of about 22 minutes, with a standard deviation of 11 minutes. Our constraint should then encompass the commute of effectively all teachers.

[39] We outline the exact process for constructing these sets in Appendix C.

[40] Primary school teachers tend to have one classroom, but secondary school teachers may have multiple. We consider the average to account for this.

[41] Santelli and Grissom (2022) find that a teacher with a 45-minute commute has a 10 p.p. greater likelihood of leaving compared to another teacher who works at the same school but experiences a 5-minute commute.

[42] Surveyed teachers do not explain what they spend their time doing, but we assume the work week hours reported do not include the time it takes to drive to their school.

observable and unobservable attributes of teachers; formally,

$$\alpha_i = \alpha_0 + \alpha_1 x_i^{\text{male}}; \qquad \beta_i = \beta_0 + \beta_1 x_i^{\text{male}} + \beta_2 x_i^{\text{NW}} + \beta_3 x_i^{\text{c}}; \qquad \gamma_i = \gamma_0 + \gamma_1 x_i^{\text{NW}},$$

where x_i^{male} equals 1 if the teacher is male, x_i^{NW} equals 1 if the teacher is non-white, and x_i^{c} equals 1 if the teacher is committed.[43] We anchor the utility teachers receive from their salary by setting $\alpha_0 = 1$, which provides meaningful interpretation for all other parameter estimates as a relative willingness to pay. Finally, the term ϕ_i is a random coefficient that enters into the utility for teaching, i.e., it is sector-specific and not school-specific. We assume that this varies by commitment type in the following way:

$$\phi_i = \begin{cases} \phi_{uc} \sim \mathcal{N}(0,1), & \text{if } x_i^c = 0 \\ \phi_c \sim \mathcal{N}(\mu_c, \sigma_c^2), & \text{otherwise.} \end{cases}$$

In the above, μ_c captures the average additional utility enjoyed by a committed teacher from working in the teaching profession. For completeness, $\varepsilon_{ij\ell t} \sim \mathcal{N}(0, \sigma_{\varepsilon,j}^2)$ captures *iid* idiosyncratic shocks to taste for working in schools.

We let salary offers differ by teaching and non-teaching sectors. The majority of states over the period we analyze mandate schools to follow a wage schedule which depends only on teaching experience.[44] Our data considers only new teachers, so among teachers that stay, the number of years of teaching experience a teacher has is perfectly collinear to their salary that period. As a result, we cannot estimate regressions and instead assume that (log) annual salary follows the distribution $\ln w_{jt} \sim \mathcal{N}(\mu_t^w, \sigma_t^{w^2})$, where teachers that stay at the same school period-to-period receive an updated wage according to $\ln w_{jt+1} = 1.01 \times \ln w_{jt}$, which captures the small growth in annual wages as outlined by most district schedules. Teachers who switch schools redraw from the distribution for $\ln w_{jt+1}$.

Non-teaching jobs are not required by law to give identical salary offers to observably different applicants. As a result, to recover predictions for these salaries, we estimate regressions of the

[43] In essence, we aggregate teachers into one of eight gender-race-commitment types. Figure B2 provides the relative composition of these types in the sample.

[44] In 2008, 18 states give a premium to teachers with a master's degree. Otherwise, the sole determinant in salary is the number of teaching years incurred.

form

$$\ln w_s = \boldsymbol{x}_i' \boldsymbol{\beta}^{\text{HC}} + \varepsilon_i^{\text{HC}}, \tag{2}$$

where $\boldsymbol{\beta}^{\text{HC}}$ denotes the vector of coefficients obtained from a two-step Heckman correction procedure and $\boldsymbol{x}_i$ is a vector of covariates that includes gender, race, an interaction of the two, education, degree type, non-teaching experience, and U.S. region of residence. Given the small number of teachers that exit the teacher labor force, we estimate the model using the final year of the sample, when the most teachers have exited from schools. Therefore, we drop t subscripts and assume a firm is defined by the same wage in the second and third periods. The small number of realizations further prevents us from estimating sector-specific coefficients or from including sector fixed effects. We present the results from this procedure in Table A5.

Unknown Non-Pecuniary Amenity

Given our assumption that some school amenities become known to teachers only *after* accepting a job offer, we next outline the way teachers form expectations. In our model, teachers care about the frequency of behavioral infractions at a school but cannot observe the true school-level rate of behavioral incidents until they begin working at that school. We measure behavioral infractions as a relative frequency of the weighted average of the 10 most common types of infractions.[45] Teachers observe whether the school extending them an offer resides in a ZIP code with below median (L) or above median (H) socioeconomic status (SES), denoted by ℓ. Given ℓ, teachers know the true distribution of (log) behavioral infractions, $\ln b_{j\ell} \sim \mathcal{N}(\mu_\ell, \sigma_\ell^2)$, but receive a noisy signal of the form $\ln \tilde{b}_{j\ell} = \ln b_{j\ell} + \nu_i$, where the error term follows a distribution according to $\nu_i \sim \mathcal{N}(0, \sigma_{\nu,i}^2)$, and where the variance flexibly allows for the same heterogeneous margins as in the preference parameters. This captures the fact that observably different teachers may be more informed or put forth more effort to learn about a school's hidden amenities before accepting an offer. We parameterize this term in the following way: $\sigma_{\nu,i} = \vartheta_0 + \vartheta_1 x_i^{\text{male}} + \vartheta_2 x_i^{\text{NW}} + \vartheta_3 x_i^{\text{c}}$. Conditional on the signal, the expected value of behavioral infractions becomes

$$\ln b_{ij\ell}^E = \big(1 - \varphi_{i\ell}\big)\mu_\ell + \varphi_{i\ell} \ln b_{j\ell} + \varphi_{i\ell}\nu_i, \quad \text{where} \quad \varphi_{i\ell} = \frac{\sigma_\ell^2}{\sigma_\ell^2 + \sigma_{\nu,i}^2}.$$

[45] These responses are provided by principals of available schools in the SASS. The list of infractions is found in Appendix C.

Note that $\varphi_{i\ell} \to 1$ as $\sigma^2_{\nu,i} \to 0$, which yields $\ln b^E_{ij\ell} \to \ln b_{j\ell}$.[46] Given this structure, we can write the *expected* utility that teacher i receives from working at school j with SES level ℓ in period t as

$$\begin{aligned}
\tilde{u}_{ij\ell t} &= \alpha_i \ln w_{j\ell t} + \underbrace{\beta_i\big(1-\varphi_{i\ell}\big)\mu_{i\ell}}_{\text{fixed effect, } \tilde{\mu}_{i\ell}} + \varphi_{i\ell}\beta_i \ln b_{j\ell} + \varphi_{i\ell}\beta_i\nu_i + \ldots + \phi_i + \varepsilon_{ij\ell t} \\
&= \alpha_i \ln w_{j\ell t} + \tilde{\mu}_{i\ell} + \varphi_{i\ell}\beta_i \ln b_{j\ell} + \ldots + \phi_i + \underbrace{\varphi_{i\ell}\beta_i\nu_i + \varepsilon_{ij\ell t}}_{\text{residual term, } \tilde{\varepsilon}_{ij\ell t}} \qquad (3) \\
&= \alpha_i \ln w_{j\ell t} + \tilde{\beta}_{i\ell} \ln b_{j\ell} + \ldots + \lambda(40 - h^E_{ij\ell t}) + \psi_1 \bar{n}_{ij\ell t} + \psi_2 \bar{n}^2_{ij\ell t} + \tilde{\mu}_{i\ell} + \phi_i + \tilde{\varepsilon}_{ij\ell t},
\end{aligned}$$

where $\tilde{\beta}_{i\ell} = \varphi_{i\ell}\beta_i$ and $\mathbb{E}[\tilde{\varepsilon}_{ij\ell t}] = 0$ by construction. Note that the fixed effect $\tilde{\mu}_{i\ell} = 0$ whenever teacher i has perfect information. Conversely, when the signal is extremely noisy, $\varphi_i \to 0$ and teachers rely only on the expected mean of school climate. In this instance, the true value of their own school's climate has no impact on their decision-making process.

Lastly, we assume that new teachers initially anticipate 40-hour work weeks at every job, meaning the λ term disappears in the expected utility during the initial match decision. Teachers learn the true value of $h_{ij\ell t}$ during the school year and update their beliefs in subsequent periods according to $h^E_{ij\ell t+1} \coloneqq \mathbb{E}[h_{ij\ell t+1} \mid h_{ij\ell t}] = h_{ij\ell t}$. In other words, teachers expect the deviation in workload to persist over time. The lack of data on hours worked at non-teaching jobs requires us to further assume that these jobs are defined by strict 40-hour work weeks.

Career Choices

In subsequent periods, teachers choose among four discrete choices: remain at their current school (j), remain in teaching but switch schools (j'), remain in the labor market but switch jobs (s), or exit the labor market (the outside option, 0). Teachers face switching costs whenever their decision represents a change compared to the previous period. We allow for the cost of switching schools (κ_j) to differ from the cost of switching sectors (κ_s).

We now define the expected utility a teacher receives from each of the remaining three alter-

[46] For the full explanation of this derivation, see Appendix C.

natives. A teacher who switches to school $j' \neq j$ expects to receive utility

$$\tilde{u}_{ij'\ell t} = \alpha_i \ln w_{j'\ell t} + \tilde{\beta}_{i\ell} \ln b_{j'\ell} + \ldots + \psi_1 \bar{n}_{ij'\ell t} + \psi_2 \bar{n}^2_{ij'\ell t} + \tilde{\mu}_{i\ell} + \phi_i - \kappa_j + \tilde{\varepsilon}_{ij'\ell t},$$

while a teacher who switches careers to work in some non-teaching sector s expects to receive $\tilde{u}_{ist} = u_{ist} = \alpha_i \ln w_{st} + \delta d_{ist} - \kappa_s + \varepsilon_{ist}$. We assume that the taste shock for non-teaching jobs follows a distribution according to $\varepsilon_{ist} \sim \mathcal{N}(0, \sigma^2_{\varepsilon,s})$, which allows the variance to differ from that of the school-specific taste shock. The signals for school j' follow the same formation process as those of school j. Due to data limitations, we cannot impose similar signals to the non-teaching sector.[47] Lastly, the outside option is normalized to 0 for every teacher and for every period, i.e., $\tilde{u}_{i0t} = u_{i0t} = 0$ for each teacher in each period.

Learning

We assume that teachers perfectly learn the unobserved amenities at schools they choose to stay at for more than one period In other words, if teachers choose to stay at their initial match for the second year, their expected indirect utility is identical to their indirect utility (Equation 1). At the school margin, teachers then face a trade-off between staying at their school and knowing its amenities fully or, conditional on receiving another offer, accepting a different school with some unknown attributes. Therefore, a teacher expects to receive $\tilde{u}_{ij\ell t+1} = u_{ij\ell t+1}$

$$\tilde{u}_{ij\ell t+1} = u_{ij\ell t+1} = \alpha_i \ln w_{j\ell t+1} + \beta_i \ln b_{j\ell} + \ldots + \lambda(40 - h_{ij\ell t}) + \ldots + \phi_i + \varepsilon_{ij\ell t+1}$$

in $t+1$ from her year t match school. We also say that ϕ_i accommodates for learning in the case where the distributions of ϕ_c and ϕ_{uc} overlap. In other words, committed types whose ϕ_i draw is particularly low may have overstated their initial intent to stay in the profession. We provide a theoretical example of this learning in Figure B3.

Probabilistic Uncertainty

Teachers also face sources of probabilistic uncertainty: termination, relocation, and job offers. During the pre-tenure period, teachers must receive an annual contract renewal from their match

[47] NCES does not survey former teachers about their quality of work outside of schools. We only recover a self-reported job title and salary. As a result, we cannot recover any reasonable level of quality or firm-level demographics the same way we recover this information for schools.

school. After they earn tenure, this uncertainty disappears (Weisberg et al., 2009). In 2008, 40 states set the probationary period to three or more years (National Council on Teacher Quality, 2008).[48] As a result, the majority of teachers in our model face this uncertainty every period.

We denote the probability of being fired by p_i^f. Similar to Behrman et al. (2016), we consider a probit model to estimate the marginal propensity to be terminated for observable groups. Formally, we estimate

$$p_i^f = \mathbb{1}_{\{x_i^{\text{tenure}}=0\}} \times \Phi\Big(\rho_0^f + \rho_1^f x_i^{\text{NTX}}\Big),$$

where $\mathbb{1}_{\{x_i^{\text{tenure}}=0\}}$ is an indicator function which takes the value of 1 when teacher i has not earned tenure. We assume teachers know this exogenous value and that it does not change over time.

Teachers also face relocation uncertainty. We distinguish between moving *within* state (p_i^{rw}) and moving *between* states (p_i^{rb}). Notably, either refers to moving beyond the initial consideration zone.[49] We assume the probabilities of being relocated each follow a probit according to

$$p_i^r = \Phi\Big(\rho_0^r + \rho_1^r x_i^{\text{NTX}}\Big),$$

where $r \in \{rw, rb\}$. From the second year onward, teachers may receive offers from teaching and non-teaching jobs. We assume each teacher receives one non-teaching job offer each year with certainty, defined by its log salary and its distance from home, $\{w_s, d_{is}\}$. On the other hand, teachers face a probability of receiving a non-teaching offer, p_i^o. We recover this exogenous probability in two steps. We first predict the number of schools teacher i applies to, $|\widehat{\mathcal{A}}_i|$, and then predict then number of schools i receives an offer from, $|\widehat{\mathcal{O}}_i|$. Then the offer probability takes the value $p_i^o = \frac{|\widehat{\mathcal{O}}_i|}{|\widehat{\mathcal{A}}_i|}$. We present the distribution of $|\widehat{\mathcal{O}}_i|$ and $|\widehat{\mathcal{A}}_i|$ in Figure B6.[50]

[48] California, Maine, Maryland, Nevada, South Carolina, Vermont, and Washington set their probation period to two years, while Hawaii, Mississippi, and North Dakota set it to one year. At the time, D.C. did not award tenure.

[49] We directly observe when a teacher moves states. We assume teachers move within a state if they switch schools and these schools are more than 28 miles away from one another.

[50] In an effort to reduce our state space size, we simplify the model by setting each teacher's likelihood of receiving a school offer to the sample average, $\rho_i^o = 0.705$.

4.3 The Teacher's Problem

Figure B7 displays the timeline of the model. We focus only on the short-run of a teacher's career, here defined as three years (i.e., $T = 3$). In the model, teacher i chooses their optimal labor path by solving a finite-horizon, discrete choice dynamic programming problem. Their a choice set in period t, $\mathcal{C}_{it}$, denotes their current collection of teaching and non-teaching job offers. Their state variables, $\boldsymbol{\theta}_{it}$, include their tenure status, their non-teaching experience, the controls outlined in Equation 2, and, whenever $t > 1$, the pecuniary and non-pecuniary amenities of their job k_{it}. Thus, teacher i solves the Bellman equation

$$V(\boldsymbol{\theta}_{it}) = \max_{k_{it} \in \mathcal{C}_{it} \cup \{0\}} \left\{ \tilde{u}_{ikt} + \beta \mathbb{E}\Big[V(\boldsymbol{\theta}_{it+1}) \mid k_{it}, \boldsymbol{\theta}_{it}\Big] \right\},$$

where β is the shared discount factor for future periods (exogenously set to 0.96) and the expectation is taken over the sources of uncertainty (termination, relocation, and future offers). Teachers can always exit the labor market and take the outside option, denoted by $\{0\}$. We establish $\mathbb{E}\Big[V(\boldsymbol{\theta}_{iT+1}) \mid k_{iT}, \boldsymbol{\theta}_{iT}\Big] = 0$, which allows us to solve the model using backwards recursion methods (as outlined in Keane, Todd and Wolpin, 2011). In practice, we could relax this assumption and instead estimate a terminal value that captures the lifetime earnings conditional on the period three choice. However, because we only track teachers during their first three years, and given the fact that a large amount of mobility still occurs after the periods we model, we prefer this assumption to the alternative.

5 Estimation and Identification

5.1 Estimation

We estimate the model parameters through simulated method of moments (SMM). We define the parameters to estimate by $\boldsymbol{\Omega}$.[51] Under the normalization $\alpha_0 = 1$, we have 22 parameters to estimate in the model.[52] We then require at least 22 moments from the true data to identify the parameters of interest. In practice, we target 73 key moments, including but not limited

[51] We define $\boldsymbol{\Omega} := \left\{\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, \beta_3, \zeta, \gamma_0, \gamma_1, \delta, \lambda, \psi_1, \psi_2, \vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3, \mu_c, \sigma_c, \kappa_j, \kappa_s, \sigma_{\varepsilon,j}, \sigma_{\varepsilon,s}\right\}$.

[52] We estimate several other parameters, like exogenous probability parameters, offline.

to the mean and variance of accepted wages each period, the fraction of teachers that make each choice in each period, and the proportion of teachers that experience a stable match. The over-identification of our model requires the use of a weighting matrix to adjust for the relative noisiness of each targeted moment.

Denote the collection of moments generated from the data by M. We simulate each teacher's optimal labor decision path for a given set of offers, exogenous probabilities for sources of uncertainty, and tenure policies. The simulated data have a corresponding set of moments, which depend on the parameter values. We denote this collection of simulated moments by $\widetilde{M}(\mathbf{\Omega})$. The SMM-estimator satisfies

$$\widehat{\mathbf{\Omega}} = \underset{\mathbf{\Omega}}{\operatorname{argmin}} \left(M - \widetilde{M}(\mathbf{\Omega})\right)' W \left(M - \widetilde{M}(\mathbf{\Omega})\right),$$

for an optimal weighting matrix W. We set this weighting matrix equal to the precision matrix of M, $W = \Sigma^{-1}$, with Σ the covariance matrix of M.[53] To improve the numerical efficiency of the parameter estimation process, we augment the SMM estimation with an adaptive simulated annealing procedure. We then complete the procedure by using the simulated annealing solution as a starting guess in a separate nonlinear program algorithm using bounds on the search space. We compute the standard errors of each parameter via 350 bootstrap repetitions of each step of the estimation process on weighted data, including the calculation of both the targeted moments and the optimal weighting matrix, as well as the full simulated annealing procedure.

5.2 Identification Strategy

We consider several different sub-samples of the data in order to separately identify some of the key preference parameters and variance terms of our model. Notably, the information frictions introduce an identification problem in the form of $\tilde{\beta}_{i\ell} = \varphi_{i\ell}\beta_i$. For teachers that face information frictions related to student behavior, we cannot immediately recover estimates for β_i and $\varphi_{i\ell}$ separately. However, our main parameters of interest, $\{\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3\}$, can be identified after first recovering preferences for teachers who we assume have full information, i.e., those with

[53] We generate Σ by replicating the data 2,000 times via bootstrap and calculating the sensitivity of the moment across the sampling procedure. In practice, we also scale Σ using ridge regularization.

$\tilde{u}_{ij1} = u_{ij1}$. Given our assumption that teachers who stay in a school subsequently receive perfect information about all hidden amenities, we can then identify the fully-known preferences for teachers in period 1 under the following identification assumption:

> **Identification Assumption 1 (ID-1):** *New teachers in period 1 have perfect information for all hidden amenities at their initial match school if they student-taught, substitute taught, or worked any other non-teaching position at that same school or district in period 0.*

With ID-1, we have a fraction of new teachers who face no information frictions in the initial period. Let x_{i1}^{info} be a binary variable indicating whether or not a teacher has perfect information in period one. In order to recover values for $\{\vartheta_0, \vartheta_1, \vartheta_2, \vartheta_3\}$, it must be the case that $\mathbf{\Omega} \perp x_i^{\text{info}}$; that is, teacher preferences are independent to occupation in the pre-period. Given that the vast majority of teachers earn an education degree, this is reasonable if we believe college students' major choice is independent to whether the School of Education requires field work in the form of student-teaching. Our next assumption guarantees that we always have a subset of teachers with undistorted preferences in each period.

> **Identification Assumption 2 (ID-2):** *Teachers that stay at the same school period-to-period retain perfect information about that school's amenities.*

This assumption is equivalent to one which states that all non-pecuniary amenities in a school are time-invariant. A violation to this assumption would be if, for example, student proficiency was high in period t and suddenly became low in period $t+1$. However, given the relative stability of neighborhood sorting and school rankings within state, especially over a short period of time, we are not concerned with this assumption being violated.

Together, these assumptions allow us to estimate the parameters of interest by guaranteeing that in each period, a subset of teachers exists who experience no information frictions for at least one school. We further identify other preference parameters through variation in the decisions to stay, switch, and exit by observably different groups of teachers. Additionally, a small fraction of teachers begin at the same school, which provides the model with agents whose differences in decisions cannot be explained by differences in amenity bundles.

6 Results

6.1 Parameter Estimation Results

Table 4 presents the estimates of the exogenous probability parameters (Panel A) as well as the 23 structural parameters of interest (Panels B–D). Again, we standardize α_0, the baseline preference for log salary, to 1 so that each parameter has a willingness-to-pay interpretation.

We recover statistically significant estimates for exogenous probability parameters. We find that each source of uncertainty has a relatively low probability of occurrence. In addition, a teacher's level of non-teaching experience impacts uncertainty directly: teachers with more non-teaching experience have a greater likelihood of termination and a lower likelihood of relocation.

Panel B presents the estimates for parameters which we allow to vary heterogeneously in at least one dimension. In line with Bobba et al. (2021), we find that male teachers are more sensitive to salary than female teachers. All types of teachers derive large levels of disutility from student misbehavior. In particular, the parameter β_i captures the proportion of salary a teacher i would be willing to pay to avoid an increase in the rate of student misbehavior by 1 unit, or approximately 28%.[54] At the baseline, teachers are then willing to give up about 3.5% of their salary to avoid a 1% increase in student misbehavior. Male teachers, non-white teachers, and committed teachers all exhibit higher elasticity to student misbehavior than other types of new teachers.

Teachers also have different preferences for working at a school whose student body is more than 50% non-white. White teachers working at a low minority school associate a switch to a high minority school with a salary decrease of 22%. On the other hand, our model predicts strong race-based matching preferences. Among non-white teachers, a move to a majority-minority school is equivalent to a pay increase of 45%.

In the final column of Panel B, we present estimates for our main parameters of interest, the noisiness of the signal for hidden amenities. We find that all teachers are largely uninformed about the amenity they cannot observe. As evident in Equation 3, teachers down-weight the true level of misbehavior to approximately zero and take the (misaligned) average as the true value. In other words, teachers without perfect information experience φ_i close to 0 and fully rely on the term $\tilde{\mu}_{i\ell}$ when considering career choices.

[54] This corresponds to an increase of misbehavior by 1 unit from the average of 3.623 at baseline.

Table 4: Parameter Estimates

Panel A: Exogenous Probability Parameters				
	Fired (p_i^f)	Between States (p_i^{rw})	Within State (p_i^{rb})	School Offer (p_i^o)
Baseline	-2.010	-1.697	-1.647	0.705
	(0.096)	(0.078)	(0.083)	(0.307)
NT Experience	0.033	-0.045	-0.008	
	(0.013)	(0.018)	(0.013)	
Panel B: Heterogeneous Preference Parameters				
	Log Salary (α_i)	Log Behavior (β_i)	Minority Students (γ_i)	Signal Noise ($\sigma_{\nu,i}$)
Baseline	1.000	-0.990	-0.221	7.927
	(0.000)	(0.389)	(0.085)	(4.576)
× Male	0.108	-0.080		1.315
	(0.131)	(0.645)		(5.858)
× Non-White		-0.382	0.673	17.51
		(0.383)	(0.096)	(4.557)
× Committed		-1.146		18.91
		(0.407)		(4.581)
Panel C: Homogeneous Preference Parameters				
	Student Proficiency (ζ)	Travel Distance (δ)	Weekly Hours (λ)	Class Size — Linear (ψ_1)
Baseline	0.266	-0.212	0.160	0.068
	(0.076)	(0.107)	(0.057)	(0.087)
	Class Size — Quadratic (ψ_2)	School Switching Cost (κ_j)	Job Switching Cost (κ_s)	
Baseline	0.000	0.235	20.08	
	(0.001)	(0.700)	(1.726)	
Panel D: Distributional Terms				
	Committed Mean (μ_c)	Committed SD (σ_c)	School Taste Shock ($\sigma_{\varepsilon,j}$)	NT Job Taste Shock ($\sigma_{\varepsilon,s}$)
Baseline	2.676	1.893	0.129	3.514
	(1.270)	(0.636)	(0.746)	(2.866)

NOTES: NT Experience refers to the number of years of non-teaching experience held by a teacher and NT Job refers to the non-teaching job sector. Heteroskedasticity-robust standard errors in Panel A are estimated via probit regression. Standard errors in Panels B, C, and D come from 350 bootstraps of the parameter estimation procedure. SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Panel C displays estimates for homogeneous parameters. Teachers associate a school whose student body is one standard deviation higher in proficiency with a salary increase of about 27%. Somewhat similarly, teachers derive disutility from increases in travel distance. Increasing a teacher's commute by 1% is equivalent to decreasing their annual salary by 0.62%.[55] We recover small between-school switching costs but find that teachers incur high sector switching costs, as in Behrman et al. (2016). The difference in the magnitude of these switching costs can help explain why school-to-school moves account for so much of total turnover.

Lastly, we find substantial differences in the distributional terms found in Panel D. Committed

[55] The coefficient can be interpreted directly as a willingness to pay to avoid an increase in one-way commuting distance by 1 mile, or 34% relative to the average.

types derive an on-average higher fixed effect from teaching relative to self-proclaimed uncommitted types, but the distributions do overlap. Additionally, we find large differences in the size of taste shocks at schools relative to those at non-teaching jobs. Notably, the shock to non-teaching jobs has a standard deviation which is over 25 times that of teaching jobs. In conjunction with the high switching costs we recover, this noisiness helps to explain why teachers are less likely to switch jobs than they are to switch schools.

6.2 Model Validity

Our model matches both moments we target directly and those we do not target reasonably well. In total, we match 73 moments from the true data. We present a selection of some important targeted moments, and the corresponding simulated value, in Table 5. We segment the moments into one of three periods, which allows us to track whether the model also captures dynamic patterns of mobility.

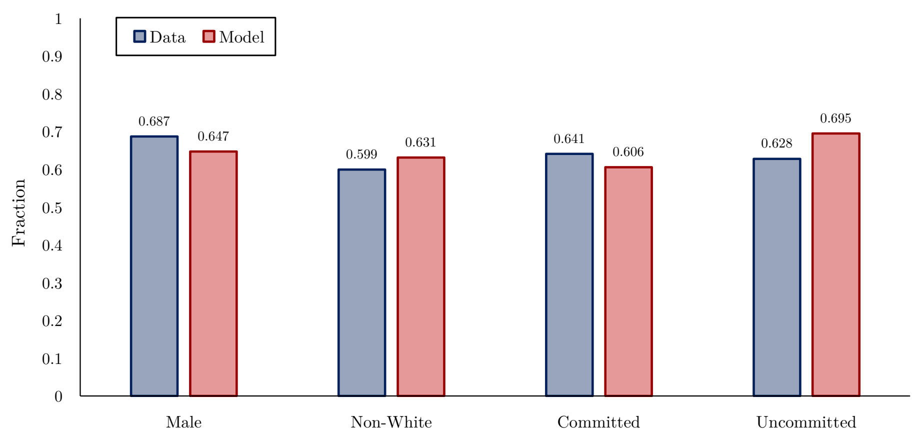
Table 5: Comparison of Targeted Moments

	Period 1		Period 2		Period 3	
	Data	Model	Data	Model	Data	Model
Panel A: Mobility Moments						
Fraction Stayers	1.000	0.998	0.775	0.793	0.742	0.747
Fraction Switchers	0.000	0.000	0.154	0.160	0.152	0.175
Fraction Leavers	0.000	0.002	0.071	0.047	0.105	0.078
Stable Match	—	—	—	—	0.637	0.633
Panel B: Amenity Averages						
Log Salary	10.47	10.45	10.19	10.28	10.14	10.26
Misbehavior Rate	3.647	3.628	3.401	3.398	3.253	3.301
Proficiency (Standardized)	0.023	0.061	0.019	0.092	0.069	0.120
Majority-Minority	0.458	0.462	0.426	0.439	0.406	0.413
Distance	2.850	2.867	3.170	3.047	3.021	3.274
Weekly Hours	55.06	54.93	50.42	51.53	47.94	48.91
Class Size	21.22	21.18	19.36	20.03	18.21	18.82

Notes: This table presents the comparison between a select few of the 73 targeted moments and the relative estimate of those moments using the parameters found under the SMM procedure. Stable match refers to whether or not a teacher stays at their initial match school for all three periods. Source: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Panel A displays relevant mobility moments, which refer to occupational choices and stable matches. The first three rows show the relative fraction of the sample choosing to either stay at their last period school (period one denotes choosing to enter teaching at all), switch to a new school, or exit the teaching profession. We match these fractions well, with slight deviations in the types of turnover. Our model tends to overestimate school-to-school and underestimate sector switching, albeit not greatly. The final row shows the fraction of teachers who experience a stable match during the three period panel, which we match with high precision. Panel B shows the model's capability to match average job amenities in every period. In each case, we match the amenity's average value nearly identically.

Figure 5: Untargeted Moments — Levels of Stable Matches



NOTES: We report the true rates of stable matches (teachers that stay at their initial match school for all three periods) in blue and the simulated levels in red. Each comparison pair conditions on the listed sub-population. SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

We validate the model using untargeted moments. We choose to explore the relative rates of stable matches among different cuts of the sample, whereas in the estimation procedure we only consider the rate for the *full* sample. In practice, we also explore the precision of moments related to accepted amenity values among untargeted sub-samples and find similar results which we do not present in this paper. Figure 5 displays the results of this validation exercise. Overall, we match these untargeted moments quite well. Among male teachers, we underestimate the true rate of stability by 4 percentage points. At the same time, we slightly overestimate (underestimate) this rate for non-white (committed) teachers. Finally, we overstate the match rate of uncommitted types by about 6 percentage points. Together, the results presented in this sec-

tion provide assurance that our model captures cross-sectional patterns of the sample well and accurately represents the underlying preferences of new teachers.

7 Counterfactual Analysis

With the model validated, we conduct six counterfactual exercises with a goal of mitigating turnover in mind. We define each counterfactual experiment, relate it to our structural model, and provide corresponding U.S. policy examples in Table 6.

Improved Non-Pecuniary Information

For this counterfactual, we simultaneously set $\vartheta_k = 0$ for each $k \in \{0,1,2,3\}$ and provide teachers with correct expectations about their workload in the first period. The first mechanism

Table 6: Counterfactual Experiments

N	Policy	Adjustment in Model	U.S. Policy Examples
I	Improved Non-Pecuniary Information	Student behavior perfectly known $\vartheta_0 = \vartheta_1 = \vartheta_2 = \vartheta_3 = 0$ First year workload known $\mathbb{E}[h_{ij\ell 1}] = 40 \rightarrow \mathbb{E}[h'_{ij\ell 1}] = h_{ij\ell 1}$	*Every Student Succeeds Act* (2015)
II	Improved Commitment Information	Least committed *committed* types switched to *uncommitted* $x_i^{c\prime} = \begin{cases} 1 & \text{if } \phi_i > \max \phi_{uc} \\ 0 & \text{otherwise} \end{cases}$	Various student-teaching college requirements
III	Universal Salary Increase	Raise all school salaries by 10% $w'_{j\ell t} = 1.1 \times w_{j\ell t}$	*American Teacher Act* (2023)
IV	Retention Bonus	Salary premium given to stayers $w'_{j\ell t+1} = (1 + 0.1 \times \mathbb{1}_{\{k_{t+1}=k_t=j\}}) \times w_{j\ell t+1}$	Flint, MI (2021); Los Angeles, CA (2023)
V	Bonus for Staff at Low-Performing Schools	Salary premium at lowest 25% of schools $w'_{j\ell t} = (1 + 0.1 \times \mathbb{1}_{\{p_z \leq p_{25}\}}) \times w_{j\ell t}$	*Teacher Loan Forgiveness Program* (1998)
VI	Relaxed Tenure Requirements	Give teachers tenure after ≤ 2 years $x_i^{\text{tenure}\prime} = \min\{2, x_i^{\text{tenure}}\}$	Indiana (2011); Ohio (2011)

Notes: This table outlines the six counterfactual experiments conducted in this paper. Column 3 describes the corresponding model adjustment. Column 4 gives examples of recent U.S. policies that relate to the counterfactual or cities or states that have enacted similar policies. The *Every Student Succeeds Act* (2015) mandates that schools publish publicly available report cards that include student proficiency level. This gave teachers access to more information regarding school quality. The examples in counterfactual II refer to Schools of Education often requiring aspiring teachers to undergo field experience (student-teaching) before their degree is conferred. These types of requirements provide incoming teachers with more on-the-job information about their fit as a teacher. Public school teachers in Flint, MI received salary bonuses of over $22,000 during Covid and the Los Angeles Unified School District unveiled a two-year retention program in 2023 that gives teachers an annual bonus of 3%. The *Teacher Loan Forgiveness Program* (1998) partially or fully forgives student loans for public school teachers in the U.S. who work at eligible Title I schools. The 118th U.S. Congress is considering passing the *American Teacher Act* (2023), which would mandate a minimum annual salary of $60,000 for public school teachers. Finally, in 2011, Indiana and Ohio both reduced their tenure requirements across the state.

allows teachers to perfectly observe student behavior *before* they match to a particular school. In reality, schools with especially high (low) levels of student misbehavior would prefer to make this information hidden (public), leading to a separating equilibrium by school behavior level. For simplicity, we assume that this policy corresponds to one in which schools become mandated to publish this information, much in the same way they publish proficiency levels now. The second mechanism corrects teachers expectations about the workload they experience at school j. We fix the $40 - h_{ij\ell t}$ utility term and instead let $h^E_{ij\ell t}$ take on the true value, $h_{ij\ell 1}$. In subsequent periods when teachers project forward, they assume the first period workload persists.

We find that fewer workers enter the teaching profession when school climate and teacher workload becomes perfectly known. In fact, the fraction of the sample matching with a school in the initial period decreases by 5.7% as a result of this policy. Those who choose not to enter in the counterfactual make this decision because their school climate is considerably worse than the average. Despite this, the rate of stable matches improves by 12% relative to the model baseline. Teachers prefer to stay at their initial match school because they no longer overestimate the quality of new schools giving them an offer, which drives down the between-school switching rate. Additionally, the fraction of teachers that ever switch schools drops by 18%. Teachers that work in either the bottom 25% or the top 25% of the student proficiency distribution are more sensitive to this policy. This points to the trade-off of the information revelation: fewer teachers enter the market, but those who do have the capacity to make well-informed, stable matches.

Improved Commitment Information

We next explore the effects of eliminating on-the-job learning by updating commitment types to match the type recovered in our estimation procedure. Our estimate for the commitment fixed effect reveals that the distribution of ϕ_i overlaps between types. We update committed types in the style outlined in Figure B3 by comparing the distributions of self-professed committed types to self-professed uncommitted types. From this, we generate a *revealed* commitment variable, $x_i^{c'}$, that takes the value 1 if $\phi_i > \max \phi_{uc}$, 0 otherwise. This process adjusts for over-commitment among new teachers by setting the bottom 73% of committed types to instead be uncommitted. Doing so changes the composition of the new teacher labor force to 18% committed, as opposed to the baseline 68% committed. This policy translates to one which provides new teachers with

more experiential learning *prior to* beginning their first year.[56]

We find that the stable match improves as a result of this policy, although the magnitude is about 40% that of the improvement found from the previous policy. Our model estimates suggest that self-professed committed types are more sensitive to student misbehavior *and* less-informed about school climate compared to uncommitted types. Therefore, when we adjust types to reflect the recovered preference for teaching, we implicitly give teachers more information about their school's climate and reduce their sensitivity to it, thereby increasing school-school switches by 3%. These results are consistent across school proficiency level.

Universal Salary Increase

A large benefit of our first two counterfactual experiments is their low implementation cost. From the perspective of the district, providing teachers with more information is effectively cost-free since schools are already required to collect and publicly release a battery of statistics related to quality. To situate the magnitude of these impacts against a higher cost policy proposed in the U.S., we consider variations in salary increases. First, we consider a flat, universal wage increase of 10% in all schools. For any school j in period t that initially offers salary w_{jt}, we instead allow them to offer $w'_{jt} = 1.1 \times w_{jt}$, with teachers continuing to earn on average 1% more each year.

According to the most recently available reports, teacher pay amounts to approximately 55% of total expenditure on public education in the United States (Irwin et al., 2022). At the same time, total expenditure on public education in the 2018–2019 school year amounted to roughly $800 billion. A back-of-the-envelope calculation reveals that the U.S. government would then need to allocate an additional $44 billion to finance this proposed policy.

As a possible solution to turnover, the universal salary increase performs worse than improved access to information. We find that stable matches increase by only 0.5%. At the same time, more teachers switch between schools since new schools are able to offer more attractive salaries. We find no substantive effects at hard-to-staff schools as a result of this policy.

Retention Bonus

Next, we consider targeted retention bonuses—again a 10% salary value—awarded to teachers that stay at school j from one year to the next. The purpose of this wage schedule is to directly incen-

[56] For example, many colleges require future graduates to student-teach at a school for part of the academic year in order to earn their degree.

tivize stable matches and minimize school-to-school switches. In the model, we adjust salaries in the following manner: $w'_{jt+1} = (1+0.1\times\mathbb{1}_{\{k_{t+1}=k_t=j\}})\times w_{jt+1}$, where $\mathbb{1}_{\{\cdot\}}$ is an indicator function that takes on the value 1 whenever teachers choose school j in both periods t and $t+1$. This schedule reduces the financial burden of the last policy by only offering the bonus to loyal teachers. In our model, approximately 80% of teachers stay at their match school in period 2, resulting in a first-year implementation cost of about $35 billion. This incentivized structure improves stable matches at a rate which is 10 times greater than that recovered under a universal salary increase. However, this policy still under-performs relative to the the information-improving policies.

Bonus for Staff at Low-Performing Schools

The last salary policy we explore targets teachers at low-performing schools, where turnover is historically highest. We again increase wages by 10%, but now only at the lowest 25% schools, measured by within-state proficiency scores. In other words, we change salaries in every period according to $w'_{jt} = (1 + 0.1 \times \mathbb{1}_{\{p_z \leq p_{25}\}}) \times w_{jt}$, where p_z is the percentile of z_j and p_{25} is the 25th percentile of a standard normal distribution. Our simulations reveal that approximately 15% of the new teacher labor force works in these schools in period 1, meaning the total cost of this plan is roughly $7 billion. Overall, this policy does not meaningfully improve stable matches or within-school switching; however, these rates improve at low-performing schools and worsen at high-performing schools. Conceptually, this policy makes offers from low-performing schools more attractive, which pulls some teachers away from better schools. Across pecuniary policies, we find that targeting low-performing schools yields better improvements to stability compared to universal salary premiums, but is outperformed by retention bonuses in every metric.

Relaxed Tenure Requirements

Our final counterfactual experiment considers the role of job protection as a mechanism to mitigate turnover. While the median teacher requires at least three years of teaching experience to earn tenure in reality, we explore the possible effects of reducing this to a universal two-year standard. In the model, this equates to shutting off termination uncertainty in period three for all teachers and preventing schools from firing teachers in the terminal period.[57] Formally, we set the new tenure requirement for each teacher to $x_i^{\text{tenure}'} = \min\{2, x_i^{\text{tenure}}\}$, and prevent any

[57] One limitation to this counterfactual is that we cannot estimate the new probit model for termination using the actual data. As a result, we assume the probability of termination remains unchanged in periods 1 and 2.

teacher with two years of experience from being fired in period three. Our results suggest that relaxing tenure requirements and thereby eliminating a source of future uncertainty for teachers leads to more favorable outcomes generally than a policy which increases salaries directly.

Discussion of Simulation Results

Table 7: Results from Counterfactual Experiments

		Improved Information		Targeted Salary Premium			Job Security
	Baseline	Amenities (I)	Commitment (II)	Universal (III)	Retention (IV)	Hard-to-Staff (V)	Relaxed Tenure (VI)
Panel A: New Teachers at All Schools							
Labor Supply (%)	99.8	94.1	99.9	99.8	99.8	99.8	99.8
		(−5.7%)	(+0.1%)	(+0%)	(+0%)	(+0%)	(+0%)
Stable Match (%)	63.3	71.2	66.4	63.6	65.5	63.5	63.5
		(+12%)	(+4.9%)	(+0.5%)	(+3.5%)	(+0.3%)	(+0.3%)
Ever Switch (%)	29.6	24.3	30.4	30.1	27.9	29.8	29.4
		(−18%)	(+2.7%)	(+1.7%)	(−5.7%)	(+0.7%)	(−0.7%)
End in Non-Teaching (%)	7.84	5.18	3.79	7.08	7.38	7.39	7.07
		(−34%)	(−52%)	(−9.7%)	(−5.9%)	(−5.7%)	(−9.8%)
Estimated Direct Cost ($B)	—	0	0	44	35.3	6.69	0.46
Panel B: New Teachers at Low-Performing Schools							
Labor Supply (%)	99.8	92.4	99.9	99.8	99.8	99.8	99.8
		(−7.4%)	(+0.1%)	(+0%)	(+0%)	(+0%)	(+0%)
Stable Match (%)	55.8	63.6	59.1	55.9	58.1	57.0	56.5
		(+14%)	(+5.9%)	(+0.2%)	(+4.1%)	(+2.0%)	(+1.3%)
Ever Switch (%)	36.1	31.0	36.6	37.2	34.3	35.7	35.7
		(−14%)	(+1.4%)	(+3.0%)	(−5.0%)	(−1.1%)	(−1.1%)
End in Non-Teaching (%)	9.16	6.44	5.26	7.96	8.71	8.44	8.90
		(−30%)	(−43%)	(−13%)	(−4.9%)	(−7.9%)	(−2.8%)
Panel C: New Teachers at High-Performing Schools							
Labor Supply (%)	99.9	92.0	100	99.9	99.9	99.9	99.9
		(−7.9%)	(+0.1%)	(+0%)	(+0%)	(+0%)	(+0%)
Stable Match (%)	66.8	75.9	69.7	66.8	68.4	66.2	67.2
		(+14%)	(+4.3%)	(+0%)	(+2.4%)	(−0.9%)	(+0.6%)
Ever Switch (%)	26.8	20.1	28.6	26.8	25.2	27.4	26.7
		(−25%)	(+6.7%)	(+0%)	(−0.7%)	(+2.2%)	(−0.4%)
End in Non-Teaching (%)	7.10	4.71	2.16	7.00	7.07	7.06	6.80
		(−34%)	(−70%)	(−1.4%)	(−0.4%)	(−0.6%)	(−4.2%)

Notes: Low- (high-) performing schools refer to those in the bottom (top) 25% of the student proficiency distribution, within-state and within-school level. In rows denoted by Stable Match, Ever Switch, and End as Non-Teaching, we condition on those that choose to enter the teaching profession in period one. The policy numbers correspond to those outlined in Table 6. Estimated direct costs represent back-of-the-envelope calculations for the first year of implementation. Source: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Table 7 presents the counterfactual experiment results. In line with our initial conjecture, we find that eliminating information frictions related to school quality improves the overall stable match rate between new teachers and schools substantially. Importantly, no other policy in our

battery of counterfactual experiments produces greater improvements to stability. A key feature of this mechanism is its low cost: schools and districts would face small time costs to include non-pecuniary information directly into publicly available reports. Given that schools already make some student information available online, the inclusion of student behavior and teacher workload would likely result in small marginal costs at most.

Costlier policies improve market conditions by small amounts at no reduction to the workforce. The largest trade-off to consider when implementing our preferred policy is the fact that the supply of new teachers will diminish as a response. While the overall rate of turnover will decrease, our results suggest that the difficulty of filling vacancies at hard-to-staff schools will likely rise. One solution to this could be to offer targeted bonuses at hard-to-staff schools *in conjunction with* information revelation. This option could have larger benefits and still maintain a lower cost compared to a universal salary increase.

8 Conclusion

Public schools in the U.S. experience markedly high levels of turnover, making them an ideal setting to study the determinants of two-sided mismatch in decentralized markets. While economists and education researchers have explored multiple competing explanations for (as well as solutions to) teacher turnover, information frictions have so far been ignored. Descriptive analyses suggest that new teachers dissolve initial school matches primarily because of non-pecuniary amenities and that they are misinformed about school characteristics that they cannot directly observe. We explore the role of these information distortions by embedding them directly into a structural model of teacher labor mobility. Our model allows for multiple sources of information frictions, including unobserved amenities, unknown workload intensity, and uncertain preferences for teaching driven by little work experience.

Our structural estimates suggest that new teachers have little information about unobserved school amenities and largely overstate their commitment to the profession. When this occurs, signals are uninformative and new teachers assume school climate matches the socioeconomic average; however, these same teachers overestimate the level of school climate available to them. This mismatch induces them to switch schools over a short period of time. Counterfactual sim-

ulations show that eliminating information frictions improves stable matches by 12%, reduces between-school switches by 18%, and reduces the fraction of former teachers by 34%. This also unintentionally reduces the labor force of new teachers by 5.7%. Correcting overcommitment in new teachers results in less sizeable effects. Compared to costly alternative policies that maintain the supply of new teachers, information revelation improves match quality most.

Together, our paper sheds light on an understudied source of instability in the market for teachers and provides credible estimates for the improvements associated with eliminating these frictions. Our results speak to the importance of this as an explanation for labor mobility and highlight the inefficiency associated with relying on salary premiums as a first-order solution.

References

Abaluck, Jason, and Giovanni Compiani. 2020. "A method to estimate discrete choice models that is robust to consumer search." National Bureau of Economic Research.

Abdulkadiroğlu, Atila, Yeon-Koo Che, Parag A Pathak, Alvin E Roth, and Olivier Tercieux. 2020. "Efficiency, justified envy, and incentives in priority-based matching." *American Economic Review: Insights*, 2(4): 425–42.

Adnot, Melinda, Thomas Dee, Veronica Katz, and James Wyckoff. 2017. "Teacher turnover, teacher quality, and student achievement in DCPS." *Educational Evaluation and Policy Analysis*, 39(1): 54–76.

Agarwal, Nikhil. 2015. "An empirical model of the medical match." *American Economic Review*, 105(7): 1939–78.

Agarwal, Nikhil, and Paulo Somaini. 2022. "Demand Analysis under Latent Choice Constraints: An Application to the US Dialysis Market."

Arcidiacono, Peter, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. 2016. "College attrition and the dynamics of information revelation." National Bureau of Economic Research.

Baldwin, Marjorie. 1991. "An asymmetric information theory of labor market discrimination." *Southern Economic Journal*, 1148–1154.

Baley, Isaac, Ana Figueiredo, and Robert Ulbricht. 2022. "Mismatch cycles." *Journal of Political Economy*, 130(11): 2943–2984.

Barnes, Gary, Edward Crowe, and Benjamin Schaefer. 2007. "The cost of teacher turnover in five school districts: A pilot study." *National Commission on Teaching and America's Future.*

Bartanen, Brendan, and Andrew Kwok. 2022. "From Interest to Entry: The Teacher Pipeline From College Application to Initial Employment."

Bates, Michael D, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin. 2022. "Teacher labor market equilibrium and student achievement." National Bureau of Economic Research.

Behrman, Jere R, Michela M Tincani, Petra E Todd, and Kenneth I Wolpin. 2016. "Teacher quality in public and private schools under a voucher system: The case of Chile." *Journal of Labor Economics*, 34(2): 319–362.

Biasi, Barbara, Chao Fu, and John Stromme. 2021. "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage." National Bureau of Economic Research.

Bobba, Matteo, Tim Ederer, Gianmarco Leon-Ciliotta, Christopher Neilson, and Marco G Nieddu. 2021. "Teacher compensation and structural inequality: Evidence from centralized teacher school choice in Perú." National Bureau of Economic Research.

Bonhomme, Stéphane, Grégory Jolivet, and Edwin Leuven. 2016. "School characteristics and teacher turnover: Assessing the role of preferences and opportunities." *The Economic Journal*, 126(594): 1342–1371.

Buchholz, Nicholas. 2022. "Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry." *The Review of Economic Studies*, 89(2): 556–591.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American economic review*, 104(9): 2633–79.

Coffman, Lucas C, John J Conlon, Clayton R Featherstone, and Judd B Kessler. 2019. "Liquidity affects job choice: Evidence from Teach for America." *The Quarterly Journal of Economics*, 134(4): 2203–2236.

Cooper, James M, and Amy Alvarado. 2006. *Preparation, recruitment, and retention of teachers.* International Institute for Educational Planning Brussels.

Dai, Xiaowu, and Michael I Jordan. 2021. "Learning strategies in decentralized matching markets under uncertain preferences." *Journal of Machine Learning Research*, 22(260): 1–50.

Dolton, Peter, and Wilbert van der Klaauw. 1999. "The turnover of teachers: A competing risks explanation." *Review of Economics and Statistics*, 81(3): 543–550.

Eckstein, Zvi, and Kenneth I Wolpin. 1990. "Estimating a market equilibrium search model from panel data on individuals." *Econometrica: Journal of the Econometric Society*, 783–808.

Ederer, Tim. 2023. "Labor market dynamics and teacher spatial sorting."

Falch, Torberg, and Bjarne Strøm. 2005. "Teacher turnover and non-pecuniary factors." *Economics of Education Review*, 24(6): 611–631.

Feng, Li. 2009. "Opportunity wages, classroom characteristics, and teacher mobility." *Southern Economic Journal*, 75(4): 1165–1190.

Golan, Limor. 2005. "Counteroffers and efficiency in labor markets with asymmetric information." *Journal of Labor Economics*, 23(2): 373–393.

Gray, Lucinda, and Soheyla Taie. 2015. "Public School Teacher Attrition and Mobility in the First Five Years: Results from the First through Fifth Waves of the 2007-08 Beginning Teacher Longitudinal Study. First Look. NCES 2015-337." *National center for education statistics.*

Hanushek, Eric A, and Steven G Rivkin. 2010. "Constrained job matching: Does teacher job search harm disadvantaged urban schools?" National Bureau of Economic Research.

Hanushek, Eric A, Steven G Rivkin, and Jeffrey C Schiman. 2016. "Dynamic effects of teacher turnover on the quality of instruction." *Economics of Education Review*, 55: 132–148.

Ingersoll, Richard M, Elizabeth Merrill, Daniel Stuckey, and Gregory Collins. 2018. "Seven trends: The transformation of the teaching force–updated October 2018."

Irwin, Véronique, Josue De La Rosa, Ke Wang, Sarah Hein, Jijun Zhang, Riley Burr, Ashley Roberts, Amy Barmer, Farrah Bullock Mann, Rita Dilig, et al. 2022. "Report on the Condition of Education 2022. NCES 2022-144." *National Center for Education Statistics.*

Johnston, Andrew C. 2020. "Preferences, selection, and the structure of teacher pay." *Available at SSRN 3532779.*

Keane, Michael P, Petra E Todd, and Kenneth I Wolpin. 2011. "The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications." In *Handbook of labor economics.* Vol. 4, 331–461. Elsevier.

Lam, Onyi, Brian Broderick, and Skye Toor. 2018. "How far Americans live from the closest hospital differs by community type."

Li, Hao, Sergei Severinov, and Yujing Xu. 2020. "Dating and Divorce."

Lindqvist, Per, Ulla Karin Nordänger, and Rickard Carlsson. 2014. "Teacher attrition the first five years–A multifaceted image." *Teaching and teacher education*, 40: 94–103.

Morgan, Andrew J, Minh Nguyen, Eric A Hanushek, Ben Ost, and Steven G Rivkin. 2023. "Attracting and Retaining Highly Effective Educators in Hard-to-Staff Schools." National Bureau of Economic Research.

National Council on Teacher Quality. 2008. "State Teacher Policy Yearbook 2008." National Council on Teacher Quality.

Niederle, Muriel, and Leeat Yariv. 2007. "Matching through decentralized markets." *Discussion Paper, Stanford University.*

OECD. 2018. *OECD Teaching and Learning International Survey (TALIS): Volume II.* Organisation for Economic Cooperation and Development.

Ronfeldt, Matthew, Susanna Loeb, and James Wyckoff. 2013. "How teacher turnover harms student achievement." *American educational research journal*, 50(1): 4–36.

Santelli, Francisco Arturo, and Jason A Grissom. 2022. "A Bad Commute: Does Travel Time to Work Predict Teacher and Leader Turnover and Other Workplace Outcomes?"

Simon, Nicole, and Susan Moore Johnson. 2015. "Teacher turnover in high-poverty schools: What we know and can do." *Teachers College Record*, 117(3): 1–36.

Sorensen, Lucy C, and Helen F Ladd. 2020. "The hidden costs of teacher turnover." *Aera Open*, 6(1): 2332858420905812.

Steinberg, Matthew P, Ruth Curran Neild, W Kyle Canuette, Sharin Park, Emily Schulman, and Melissa Wright. 2018. "Teacher Mobility in the School District of Philadelphia, 2009-10 through 2015-16." *Philadelphia Education Research Consortium.*

Stinebrickner, Todd R. 2001. "A dynamic model of teacher labor supply." *Journal of Labor Economics*, 19(1): 196–230.

Stinebrickner, Todd R. 2002. "An analysis of occupational change and departure from the labor force: Evidence of the reasons that teachers leave." *Journal of Human Resources*, 192–216.

Tsai, Yi-Lin, and Elisabeth Honka. 2021. "Informational and noninformational advertising content." *Marketing Science*, 40(6): 1030–1058.

Van den Berg, Gerard J, and Geert Ridder. 1998. "An empirical equilibrium search model of the labor market." *Econometrica*, 1183–1221.

Walker, Tim. 2022. "Survey: Alarming Number of Educators May Soon Leave the Profession."

Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, David Keeling, Joan Schunck, Ann Palcisco, and Kelli Morgan. 2009. "The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness." *New Teacher Project.*

Wiswall, Matthew. 2007. "Licensing and occupational sorting in the market for teachers." *Unpublished manuscript, Department of Economics, New York University.*

Zimmerman, Alex. 2017. "40 percent of teachers were gone from struggling New York City schools after two years, union data show."

Online Appendix

A Additional Tables

Table A1: Evidence for Randomness in Period 3 Missing Status

	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.027	0.026	0.026	0.028	0.028	0.031
	(0.023)	(0.023)	(0.023)	(0.025)	(0.025)	(0.024)
Black		-0.011	-0.006	-0.008	-0.008	-0.017
		(0.017)	(0.017)	(0.018)	(0.018)	(0.019)
Hispanic		0.060	0.060	0.054	0.055	0.051
		(0.048)	(0.049)	(0.049)	(0.049)	(0.046)
Other Race		-0.006	-0.007	-0.009	-0.008	-0.020
		(0.022)	(0.023)	(0.024)	(0.024)	(0.024)
Age			-0.001	0.000	0.000	-0.000
			(0.008)	(0.008)	(0.008)	(0.008)
Age^2			-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Union				-0.008	-0.009	-0.003
				(0.019)	(0.018)	(0.018)
Low Quality					-0.007	-0.007
					(0.022)	(0.022)
Medium Quality					-0.020	-0.017
					(0.017)	(0.018)
Same School (Period 2)						-0.112
						(0.063)
Different School (Period 2)						-0.086
						(0.064)
Different Job (Period 2)						-0.051
						(0.065)
Period 1 School FEs	✗	✗	✗	✓	✓	✓
Period 2 School FEs	✗	✗	✗	✗	✗	✓
Observations	1,870	1,870	1,870	1,870	1,870	1,870

NOTES: We report results from a regression of $missing_i$ on a specified vector of controls, where $missing_i$ equals 1 if teacher i did not fill out the survey in period T, 0 otherwise. Each regression uses analysis weights provided by the National Center for Education Statistics (NCES). We report heteroskedastic-robust standard errors in parentheses below each point estimate. The fixed effects in each period refer to school level type (elementary, middle, high) and rurality level. In every specification, none of the fixed effects yield estimates significant at the 5% level. SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Table A2: OLS Results for the Likelihood of Commitment

	Probability of Commitment			
	(1)	(2)	(3)	(4)
Female	0.022	0.020	0.014	0.011
	(0.042)	(0.041)	(0.039)	(0.039)
Non-White	0.015	0.034	0.023	0.042
	(0.050)	(0.051)	(0.049)	(0.053)
Age	0.003	0.002	0.003	0.003
	(0.004)	(0.003)	(0.004)	(0.004)
Rural		0.049	0.032	0.050
		(0.047)	(0.047)	(0.050)
Urban		-0.047	-0.032	-0.015
		(0.058)	(0.056)	(0.058)
Master's Degree			-0.064	-0.055
			(0.050)	(0.055)
State Certification			0.001	0.003
			(0.043)	(0.044)
Education Degree			0.113**	0.090*
			(0.048)	(0.048)
Non-Teaching Experience			0.001	0.002
			(0.005)	(0.005)
Low Quality College			0.113	0.112
			(0.070)	(0.072)
Middle Quality College			0.077	0.063
			(0.054)	(0.055)
State Fixed Effects	✗	✗	✗	✓
Observations	1,740	1,740	1,740	1,740

NOTES: The above table displays coefficients and heteroskedastic-robust standard errors from regressions of a binary indicator for commitment on a host of observable attributes of teachers. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Table A3: Applications and Offers for Early Career Teachers

	Log Number of Applications			Log Number of Offers		
	(1)	(2)	(3)	(4)	(5)	(6)
Age	−0.053	−0.079	−0.097	−0.033	−0.086	−0.048
	(0.157)	(0.156)	(0.153)	(0.074)	(0.251)	(0.075)
Male	0.261**	0.312***	0.134	−0.090*	−0.061	−0.139**
	(0.107)	(0.108)	(0.110)	(0.051)	(0.055)	(0.054)
Minority	−0.530***	−0.454***	−0.325**	0.094	0.125*	0.088
	(0.128)	(0.129)	(0.131)	(0.061)	(0.066)	(0.065)
Married	−0.035	−0.109	0.059	0.048	0.031	0.064
	(0.098)	(0.100)	(0.098)	(0.047)	(0.051)	(0.048)
Low HH Income	−0.203**	−0.188**	−0.231**	−0.045	−0.029	−0.062
	(0.096)	(0.096)	(0.093)	(0.046)	(0.049)	(0.046)
High HH Income	0.219	0.265*	0.246	−0.143*	−0.148*	−0.141*
	(0.158)	(0.158)	(0.153)	(0.075)	(0.078)	(0.075)
Education Major	–	0.175*	0.043	–	0.063	0.097*
		(0.106)	(0.109)		(0.054)	(0.054)
State Certified	–	0.271**	0.176	–	0.010	−0.037
		(0.129)	(0.130)		(0.068)	(0.064)
Elementary	–	–	−0.013	–	–	−0.163**
			(0.122)			(0.060)
Middle	–	–	0.177	–	–	−0.077
			(0.131)			(0.064)
Intercept	1.553***	1.238**	2.086**	0.653**	1.015***	0.900**
	(0.167)	(0.191)	(0.222)	(0.079)	(0.387)	(0.110)
Region Fixed Effects	✗	✗	✓	✗	✗	✓
Observations	950	950	940	950	950	940

NOTES: Age is an indicator for whether an individual's age is above the average in the sample. We present heteroskedasticity-robust standard errors in parentheses beneath each point estimate. Sample sizes are rounded to the nearest 10 in compliance with guidelines issued by the National Center for Education Statistics (NCES). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. SOURCE: U.S. Department of Education, National Center for Education Statistics, Baccalaureate and Beyond Longitudinal Study (B&B), 1993-2003.

Table A4: Explanatory Variable Comparison for Two Education Datasets

	B&B Sample		BTLS Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.749	0.434	0.758	0.428
Non-White	0.149	0.356	0.209	0.407
Age	26.42	12.25	28.98	8.419
Married	0.332	0.471	0.445	0.497
High HH Income	0.099	0.299	0.110	0.313
Education Major	0.610	0.488	0.724	0.447
State Certified	0.794	0.404	0.667	0.471
Elementary	0.504	0.500	0.454	0.498
Middle	0.247	0.431	0.177	0.382
High	0.249	0.433	0.304	0.460
Log Applications	1.377	1.368	1.176	0.518
Log Offers	0.594	0.650	0.705	0.182
Observations	950		1,740	

Notes: This table displays summary statistics for key explanatory variables used to generate application set sizes and offer set sizes. The first two columns display the spread for variables using the Baccalaureate and Beyond (B&B) sample; the next two columns display the same using the Beginning Teacher Longitudinal Study (BTLS) data. "Log Applications" and "Log Offers" respectively refer to the log number of applications sent out and the log number of offers received in the teacher labor market. Sources: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10; U.S. Department of Education, National Center for Education Statistics, Baccalaureate and Beyond Longitudinal Study (B&B), 1993-2003.

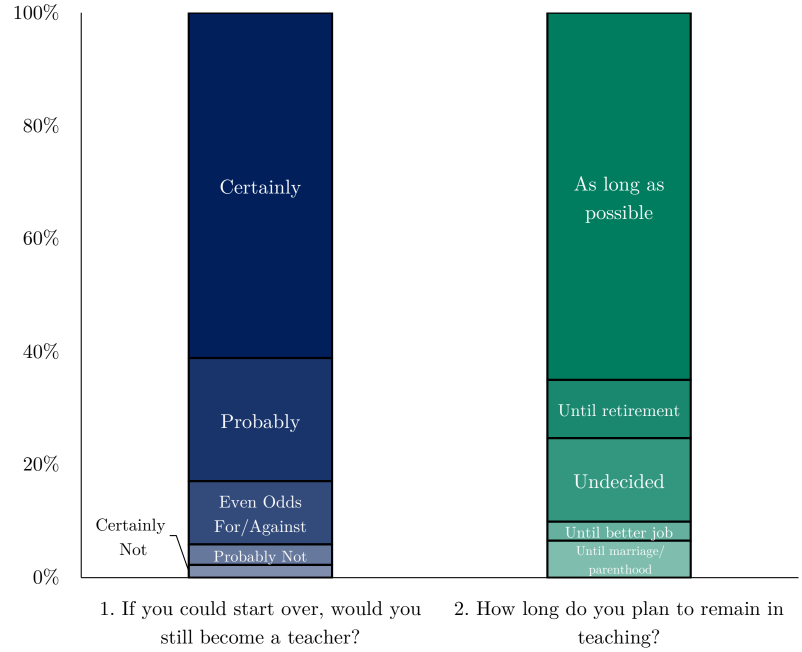
Table A5: Selection-Corrected Non-Teaching Wage Regression Coefficients

	Log Wage	Selection into Non-Teaching
	(1)	(2)
Female	−0.278*	0.072
	(0.158)	(0.102)
Non-White	0.177	0.123
	(0.292)	(0.205)
Female×Non-White	0.079	−0.321
	(0.343)	(0.255)
Master's	−0.039	0.119
	(0.170)	(0.110)
Education Degree	−0.155	−0.093
	(0.142)	(0.107)
Non-Teaching Experience	0.033***	0.014*
	(0.009)	(0.008)
Midwest Region	0.021	−0.034
	(0.246)	(0.148)
South Region	0.419*	−0.173
	(0.217)	(0.145)
West Region	0.006	−0.039
	(0.206)	(0.147)
Fired from Teaching		1.348***
		(0.202)
Married		−0.163*
		(0.094)
Number of Children		0.031
		(0.045)
Constant	10.141***	−1.310***
	(0.397)	(0.161)
Inverse Mills Ratio	−0.152	
	(0.187)	
Observations	160	1740

NOTES: Coefficients in column 1 display estimates from regressions of the determinants of log wage in non-teaching jobs, using a Heckman selection correction two-step procedure. The coefficients from the estimated selection process—deciding whether to enter the non-teaching sector—are presented in column 2. Robust standard errors are displayed in parentheses. Region indicators are relative to the omitted Northeast Region. Observation counts are rounded to the nearest 10 to preserve anonymity. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

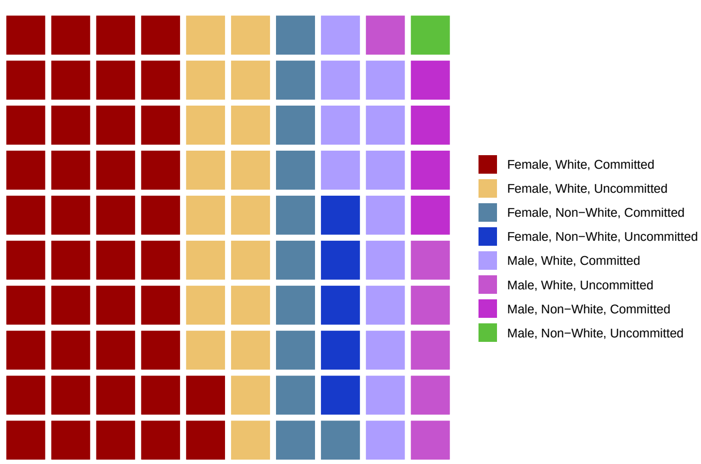
B Additional Figures

Figure B1: Response Distribution for Questions on Commitment Level



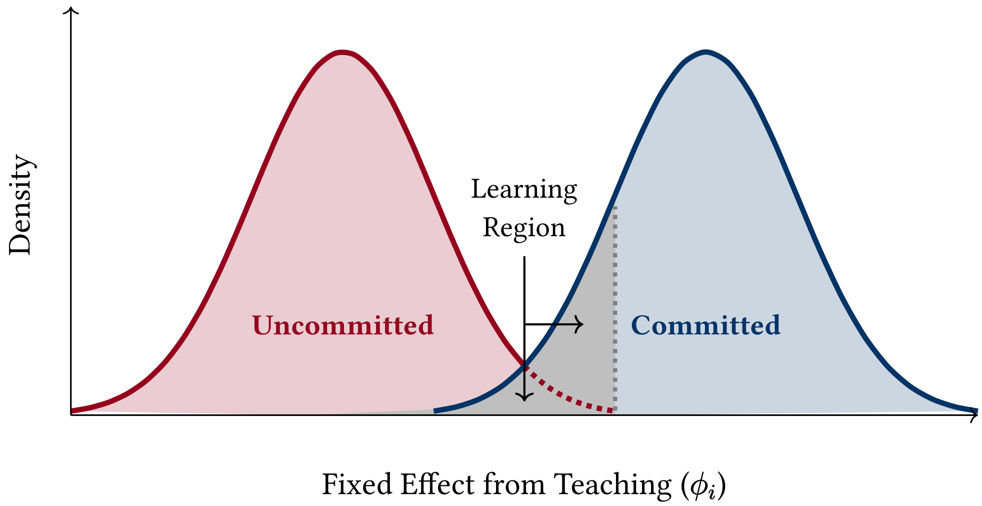
Notes: We report weighted percentages obtained using final sample weights provided by NCES. Teachers are denoted as *committed* if they answer one of the top two responses in question 1 *and* they answer one of the top two responses in question 2. Source: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Figure B2: Heterogeneity in Teacher Type by Observable and Unobservable Characteristics



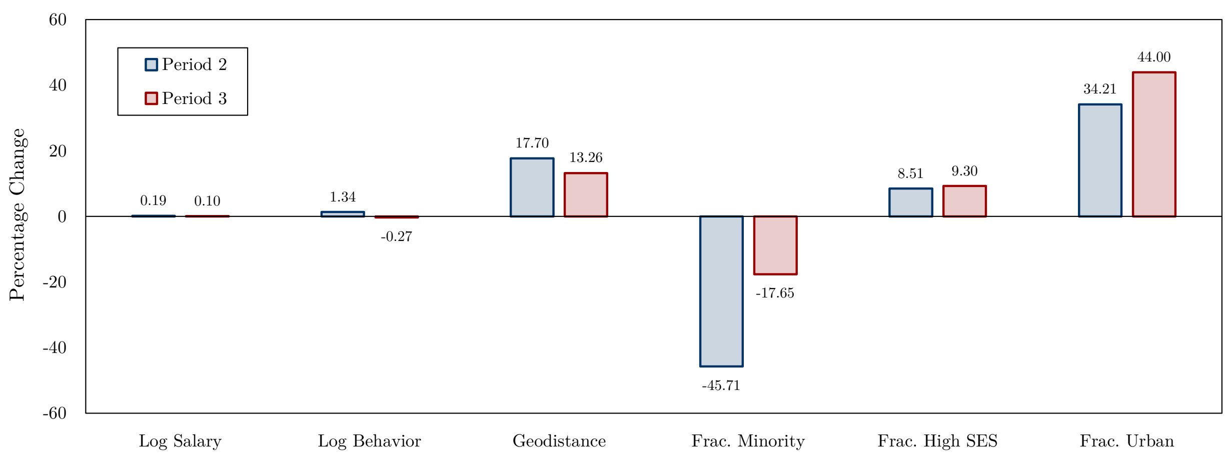
Notes: Each grid represents 1% of the total sample of 1,740 new teachers. Rounding errors are inconsistently applied to preserve a total sum of 100%. The final group is shown to be 1% of the sample despite actually representing 1.55%. Otherwise, all groups display the correct (rounded) proportion of the total sample they represent. Source: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Figure B3: The Possibility for Learning by Commitment Type



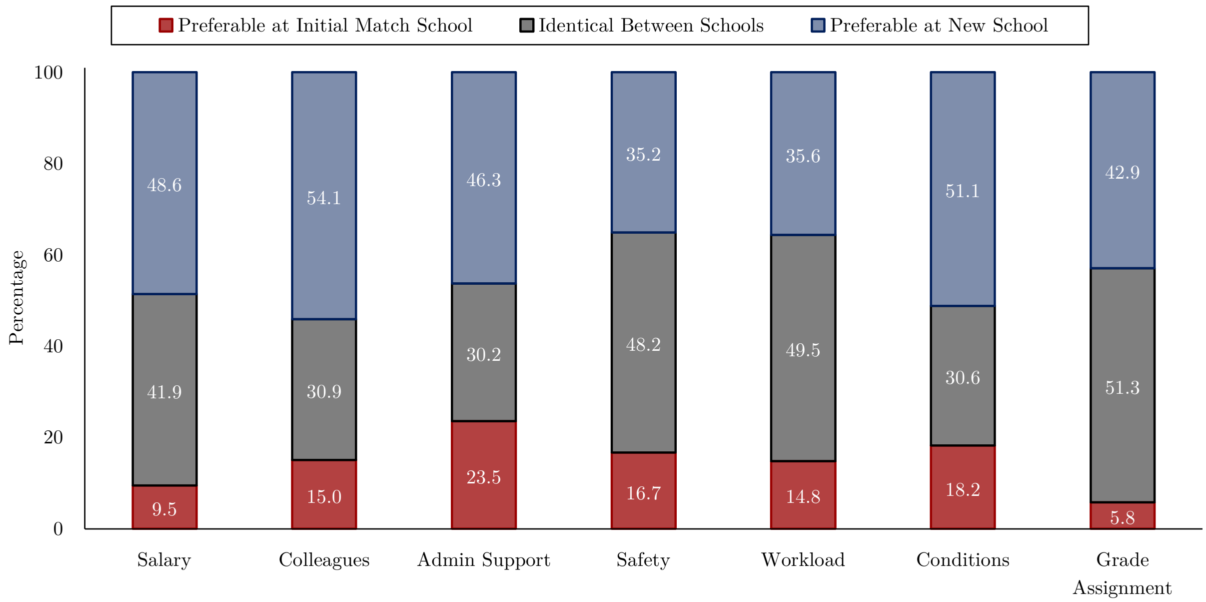
NOTES: The above figure plots the distribution of two theoretical sets of teachers, one group committed (in blue) and the other uncommitted (in red). The gray region captures the difference in the *estimated* distributions and the *stated* type from survey responses. These teachers learn that their type differs from their *ex ante* stated one through work experience.

Figure B4: Changes in School Characteristics Among Switchers



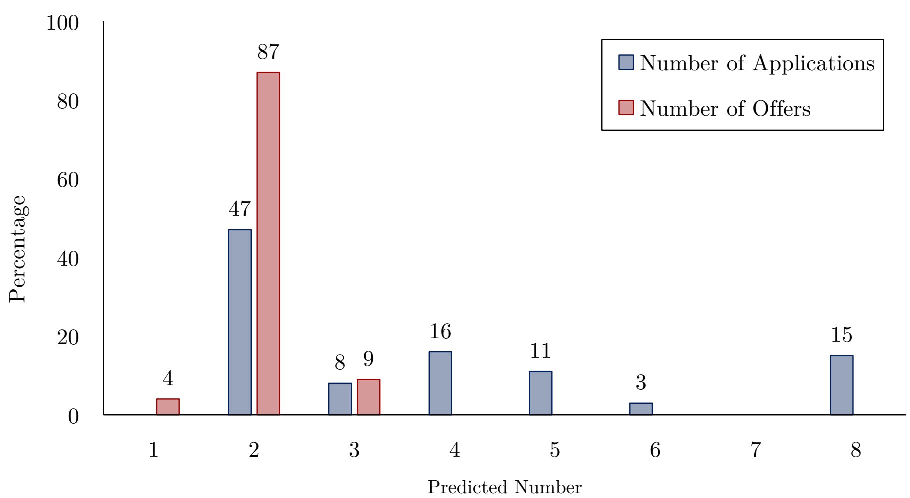
NOTES: The figure displays the relative changes in observable amenities of public schools for teachers who move either in period 2 or period 3. For simplicity, we pool these teachers and do not distinguish between the period they switch; as a result, these statistics include teachers that may remain at a school between two periods. SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Figure B5: Changes in School Characteristics Among Switchers



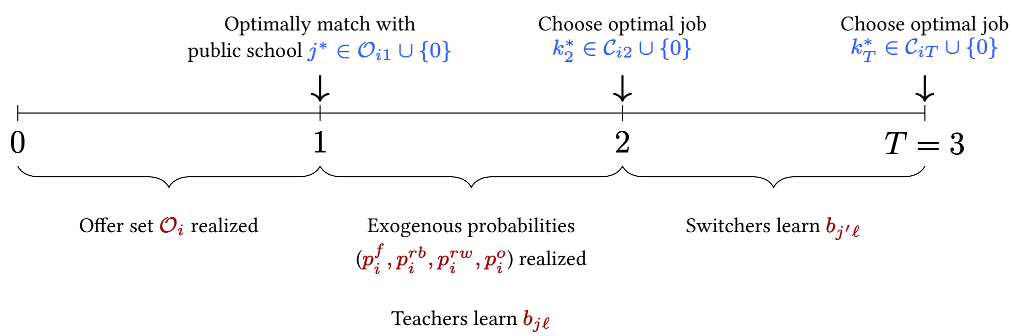
Notes: The figure displays responses from period two switchers to survey questions directly comparing attributes of their current school to their initial match school. NCES does not include similar questions in the third wave of the BTLS, meaning we cannot further compare the relative quality of schools between periods two and three among switchers. Source: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10.

Figure B6: Distribution of Predicted Applications and Offers



Notes: The figure displays the proportion of teachers in the sample that have each listed number of applications and offers. We fit regressions using the B&B and then predict the above values for the BTLS sample. For each outcome $y_i \in \{|\mathcal{A}_i|, |\mathcal{O}_i|\}$ we estimate $\ln y_i = \boldsymbol{\beta}'\boldsymbol{x}_i + \varepsilon_i$ and then exponentiate the predicted value. This procedure yields a minimum offer probability of 20% and a maximum offer probability of 100% across the sample. Source: U.S. Department of Education, National Center for Education Statistics, Beginning Teacher Longitudinal Study (BTLS), 2007-08, 2008-09, 2009-10. Source: U.S. Department of Education, National Center for Education Statistics, Baccalaureate and Beyond Longitudinal Study (B&B), 1993-2003.

Figure B7: The Decision-Making Process for Early Career Teachers



NOTES: We write exogenous variables in red and endogenous choice variables in blue. We assume that the exogenous probabilities realized in period 2 continue to apply in period 3; that is, these probabilities do not vary by time. Offers in period 2 do not predict offers in period 3.

C Detailed Procedures

C.1 Populating Period 1 Choice Sets

We implement the following procedure to construct teacher i's offer set $\mathcal{O}_i$:

1. Construct a matrix of *viable* public schools and their characteristics, where a school is viable if it belongs to the same state and contains the same grade level as the realized match school. This may include charters and non-traditional schools. Call this matrix J_i.

2. Fix the consideration radius r_i, dependent on teacher i's rurality. Compute the distance, d_{ij}, between each school $j \in J_i$ and the imputed home coordinate of teacher i. Construct $\tilde{J}_i \subseteq J_i$ where every element $j \in \tilde{J}_i$ satisfies the condition $d_{ij} \leq r_i$.

3. Randomly generate k integers $z \in \{1, ..., k\}$ among the k remaining schools. Compute the mode of these integers, denoted z^*. Include school z^* in the choice set. Ties are broken by taking the minimum of multiple modes. We set the choice set size to 2, so this completes the procedure.

C.2 Construction of Log Behavioral Infractions

We detail here how we construct the school-level measure of misbehavior, $\ln b_{j\ell}$. The School and Staffing Survey includes in the principal survey the question, "To the best of your knowledge, how often do the following types of problems occur at this school?" There are 13 listed incident types. The 10 most common infractions are as follows: student bullying, physical conflicts among students, student acts of disrespect towards teachers, robbery or theft, student verbal abuse of teachers, vandalism, student use of illegal drugs, student racial tensions, student use of alcohol, and student possession of weapons. There are five responses a principal can give for each infraction:

1. Happens daily

2. Happens at least once a week

3. Happens at least once a month

4. Happens on occasion

5. Never happens

In an effort to transform these responses into a meaningful measure of the *level* of misbehavior, we loosely map each response to the corresponding number of days in a traditional school year, with a total of 180 days. The value of any one of the k infraction measures, denoted B^k, then takes a minimum value of 1 and a maximum value of 180. We assign the value of "Happens on occasion" to 5, "Happens at least once a month" to 10, and "Happens at least once a week" to 40.

We then restrict the measure to include only the top 10 most frequent infraction types and order them accordingly. The simple weighted average takes the following form:

$$b_{j\ell} = \sum_{k=1}^{10} \frac{(10-k+1)}{10} B^k_{j\ell}.$$

We lastly transform the variable into $\ln b_{j\ell} \sim \mathcal{N}(\mu_\ell, \sigma^2_\ell)$ to adjust for the fact that $b_{j\ell}$ follows a log-normal distribution.

C.3 Derivation of Signal Expectation

Consider two random variables X and Y whose joint distribution follows

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma^2_X & 0 \\ 0 & \sigma^2_Y \end{bmatrix} \right).$$

This implies that

$$X \sim \mathcal{N}\left(\mu_X, \sigma^2_X\right)$$

and

$$X + Y \sim \mathcal{N}\left(\mu_X + \mu_Y, \sigma^2_X + \sigma^2_Y\right).$$

Using the fact that $\text{Cov}(X, X+Y) = \text{Var}(X)$, we have

$$\begin{pmatrix} X \\ X+Y \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_X + \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma^2_X & \sigma^2_X \\ \sigma^2_X & \sigma^2_X + \sigma^2_Y \end{bmatrix} \right).$$

Then $X \mid (X+Y) \sim \mathcal{N}(\mu, \Sigma)$, with

$$\mu = \mu_X + \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2}(X + Y - \mu_X - \mu_Y).$$

Substituting $\ln b_{j\ell} \sim \mathcal{N}(\mu_\ell, \sigma_\ell^2)$ for $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $\nu_i \sim \mathcal{N}(0, \sigma_{\nu,i}^2)$ for $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, we have for some signal $\tilde{b}_{ij\ell}$ the *conditional expected behavioral infractions* as

$$\begin{aligned}
\ln b_{ij\ell}^E &= \mu_\ell + \frac{\sigma_\ell^2}{\sigma_\ell^2 + \sigma_{\varepsilon,i}^2}(\ln b_{j\ell} + \nu_i) - \frac{\sigma_\ell^2}{\sigma_\ell^2 + \sigma_{\nu,i}^2}(\mu_\ell + 0) \\
&= \frac{\sigma_{\nu,i}^2}{\sigma_\ell^2 + \sigma_{\nu,i}^2}\mu_\ell + \frac{\sigma_\ell^2}{\sigma_\ell^2 + \sigma_{\nu,i}^2}\ln b_{j\ell} + \frac{\sigma_\ell^2}{\sigma_\ell^2 + \sigma_{\nu,i}^2}\nu_i \\
&= \text{expected mean} + \frac{\sigma_\ell^2}{\sigma_\ell^2 + \sigma_{\nu,i}^2}\ln b_{j\ell} + \text{residual}.
\end{aligned}$$

Note that the expected mean becomes 0 if either $\mu_\ell = 0$ or $\sigma_{\nu,i}^2 = 0$.